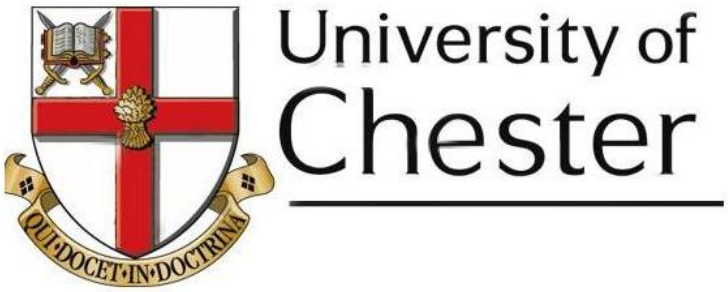

An Investigation into the Analysis of Epidemiological Models

Gerard John Coleman

October 14, 2015

Abstract

This master's dissertation concentrates on how epidemics occur, and how we can predict the likely outcomes. We use mathematical models to analyse how different population groups interact, and how certain parameters can be pivotal in accurately predicting epidemics. Throughout this dissertation we will focus on work done in the book by Murray [12] and the paper by Kermack and McKendrick [9]. We study the Human Immunodeficiency Virus (HIV), and discuss a population model which predicts how the virus behaves in a homosexual population. Then we then formulate a model of the biological make up of the disease. This enables us to evaluate the potential effects drug therapy can have in reducing the evolution of HIV into AIDS, and in reducing the spread of HIV within the population. We will evaluate how numerical approximations can provide an insight into how an epidemic may develop when analytic solutions cannot be obtained. Computed approximations will then be compared to explain the likely outcomes when an epidemic occurs.

Dissertation Supervisor: Dr Fotini Karakatsani
Title: Lecturer in Mathematics

Acknowledgments

I would like to take this opportunity to thank certain people who have made the completion of this dissertation possible. Firstly, to my supervisor Dr Fotini Karakatsani, for her knowledge and guidance. Secondly, to all staff members in the Mathematics department, especially Dr Joe Gildea and Dr Jason Roberts who have made me feel welcome at Chester University from my first day. Thirdly, I would like to thank my fellow classmates, in particular Isaac and Ruth for their support throughout the year, and to Sara for her special enthusiasm for maths and for her willingness to share food whilst working. Finally, I would like to thank my family for their support and love over the last 12 months.

Contents

Chapter 1

The SIR model

1 Introduction to Epidemic Models

Epidemiological mathematical models allow us to predict how a disease will behave once it has been introduced into a population. A suitable model will predict the evolution of the epidemic and enable authorities to implement strategies to help curb the spread and ultimately eradicate the epidemic. When creating an epidemic model, two main assumptions are made,

- Diseases do not occur at random.
- Diseases have preventative factors that can be identified within a population.

Before we can start to construct a model, it is first necessary to identify the key factors that affect the model. Firstly, it is important to identify the population who is most at risk of disease, this population may be restricted to certain age groups or sex's. Secondly, we need to identify the geographical spread of the disease; some diseases have outbreaks all over the world whilst others are confined to isolated areas. For example, the Human Immunodeficiency Virus (HIV) has spread to almost every country on earth.

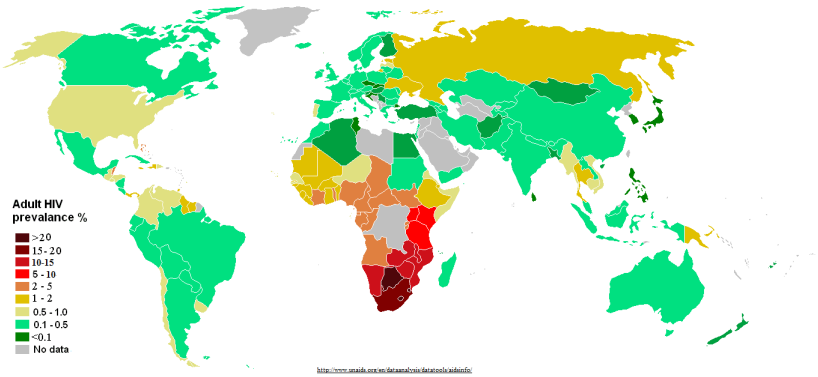

Figure 1.1: HIV worldwide spread. [19]

However, as we can see from Figure 1.1, there are concentrated areas where the prevalence of HIV exceeds 20%, in such areas the chances of an epidemic occurring is much greater than in areas with

a lower prevalence. It is therefore important to try and establish why the disease is so prevalent in areas such as southern Africa. We can do this by examining data on factors such as when does the epidemic break out, for example it may be a seasonal problem or weather dependent. We may also want to evaluate how the disease is transmitted, as this may provide reasons to why certain area's have a higher percentage of people infected with the disease. Finally, it is useful to establish which, if any, measures have already been put in place to try and combat the spread of the disease. This single factor alone may account for why certain countries have a really low prevalence, and others do not. Measures such as ensuring all children have access to education, can be essential in reducing the proportion of people who catch certain diseases, and in turn reducing the likelihood of a epidemic.

For a model to be useful, it must have the ability to adapt to new situations, be transparent, so that we can deduce how a change in a single variable may alter the overall behavior of the problem, and last but not least, the model must be accurate when compared to the collected data. Sometimes the creation of a model can be extremely difficult due to a lack of complete data to work with. Without accurate data it is very hard to deduce parameter values which are essential for creating an accurate model.

2 The SIR Model

When modeling a population of individuals we are interested in not only the size of the overall population but the size of different population groups. As it is not feasible to attempt to model the number of virus particles within the population, we instead categorise individuals into subgroups depending on their infection status. In an SIR model individuals are categorised as either

- Susceptible, S, who have not been exposed to the pathogen.
- Infected, I, who have been colonised by the pathogen.
- Recovered, R, who have fought of the infection and recovered or died due to the infection.

When there is the absence of demography within a population the standard SIR model is described as follows [9],

$$\frac{dS}{dt} = -rSI, \tag{1.1}$$

$$\frac{dI}{dt} = rSI - aI, \tag{1.2}$$

$$\frac{dR}{dt} = aI, \tag{1.3}$$

where a represents the removal rate of invectives, whilst r is the infection rate. The removal rate of invectives, is the product of the contact rate of susceptibles and invectives, and the transmission probability. Since the population has no demography it's clear that,

$$S(t) + I(t) + R(t) = N, \tag{1.4}$$

where N represents the total population. The relationship between the different populations depends on parameters a and r.

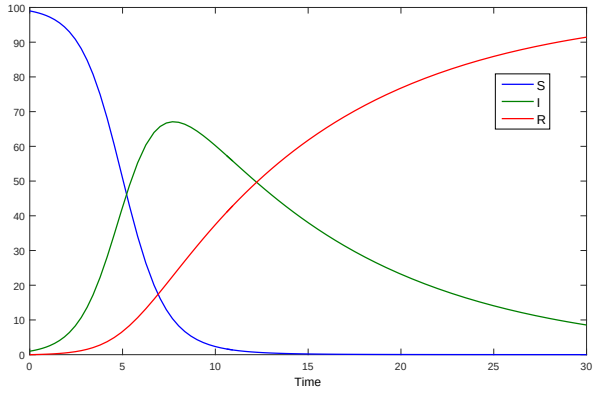

Figure 1.2: The relationship between the different population groups in the SIR model with $a = 0.1$, $r = 0.6$, $S(0) = 99$,$I(0) = 1$ and $R(0) = 0$.

Figure 1.2 illustrates that, when initial conditions are known and the parameters are estimated, the SIR model can determine whether or not an epidemic will occur. In the example used to construct Figure 1.2 the epidemic occurs around week 12. To deduce whether an epidemic may occur within a population, the basic reproduction number B_0 is a useful parameter. B_0 is the average number of secondary cases of infection arising from an average primary case within a susceptible population, hence it can be defined as $B_0 = \frac{Sr}{a}$. In an entirely susceptible population $S = 1$, hence,

- A primary infected individual spends on average $\dfrac{1}{a}$ units of time infectious.
- Per unit of time the primary infected individual contacts $a \times S = a \times 1$ individuals.
- Per unit of time these contacts result in $r \times a \times 1$ secondary infections.
- Over the period of time the primary infected individual is infected, there are resulting,

$$r \times a \times 1 \times \frac{1}{a} = \frac{r}{a}$$

  secondary infections.

- Hence,

$$B_0 = \frac{r}{a}.$$

Kermack and McKendrick's report on the mathematical theory of epidemics shows that if the proportion of Susceptibles is reduced below $\dfrac{1}{B_0}$ then the infection dies out, see [9]. Hence, an epidemic occurs when the amount of new cases arising per period of time is greater than the number of underlying cases expected in the same period of time. We can now show how we deduce this relationship, given the following initial conditions

$$S(0) = S_0 > 0, \qquad I(0) = I_0 > 0, \qquad R(0) = 0. \tag{1.5}$$

Substituting the initial conditions into equation (1.2) to get,

$$\left[\frac{dI}{dt}\right]_{t=0} = I_0(rS_0 - a). \tag{1.6}$$

Then by letting $p = \frac{a}{r}$, the change in the infective population at $t = 0$ see (1.2), is greater than zero if $S_0 > p$, and less than zero if $S_0 < p$. Indeed,

$$\left[\frac{dI}{dt}\right]_{t=0} = I_0(rS_0 - a) \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases} \quad \text{if} \quad S_0 = \begin{cases} > p \\ = p \\ < p \end{cases}.$$

As (1.1) shows S is a decreasing population for all t, which implies that $S_0 \geq S$. Therefore, we can see that if $S_0 < \frac{a}{r}$ then $S < \frac{a}{r}$ for all $t > 0$, then we can deduce the following

$$\frac{dI}{dt} = I(rS - a) \leq 0, \quad \text{for all} \quad t \geq 0. \tag{1.7}$$

Moreover, let ϕ defined as

$$\phi(t) = I(t)e^{-(rS_0 - a)t}, \quad t \geq 0. \tag{1.8}$$

In view of (1.7) and by using the fact that $S < S_0 < \frac{a}{r}$ we can see that

$$\phi'(t) = \left(\frac{dI}{dt} - (rS_0 - a)I\right) e^{-(rS_0 - a)t} \leq 0. \tag{1.9}$$

Thus, ϕ is decreasing, that is $\phi(t) \leq \phi(0), \forall t > 0$. We conclude that

$$I(t) \leq I_0 e^{(rS_0 - a)t} \tag{1.10}$$

Since $\lim_{t\to\infty} e^{(rS_0 - a)t} = 0$, we obtain that $\lim_{t\to\infty} I(t) = 0$, therefore the infection will die out, and thus an epidemic will not occur. On the other hand if $S_0 < \frac{a}{r}$ then $I(t)$ does initially increase and so an epidemic will occur. From (1.7) we can see that if $S_0 = p$ then then $\left[\frac{dI}{dt}\right]_{t=0} = 0$ hence the infective population is stationary. Eliminating time t from (1.1) and (1.2) to obtain,

$$\frac{dI}{dS} = \frac{rSI - aI}{-rSI} = -1 + \frac{aI}{rSI} = -1 + \frac{a}{rS} = -1 + \frac{p}{S}$$

i.e.,

$$dI = \left(-1 + \frac{p}{S}\right) dS.$$

We can now integrate both sides,

$$\int dI = \int \left(-1 + \frac{p}{S}\right) dS$$

to get

$$I(t) = -S(t) + p\ln(S(t)) + c, \tag{1.11}$$

where c is a constant. By applying the initial conditions to (1.11) we obtain,

$$c = I_0 + S_0 - p\ln(S_0). \tag{1.12}$$

According to (1.4) we can conclude that there are only two independent equations in the SIR model, namely the equations(1.1) and (1.2). From equation (1.4) we can also see that,

$$S + I \leq N,$$

therefore we can deduce the following conditions,

$$S(t) > \frac{p}{S_0} \quad \Rightarrow \quad \frac{dI}{dt} > 0$$

likewise,

$$S(t) < \frac{p}{S_0} \quad \Rightarrow \quad \frac{dI}{dt} < 0.$$

Since B_0 is derived as,

$$B_0 = \frac{S_0}{p}.$$

If the basic reproductive rate is greater than one, then an epidemic occurs.

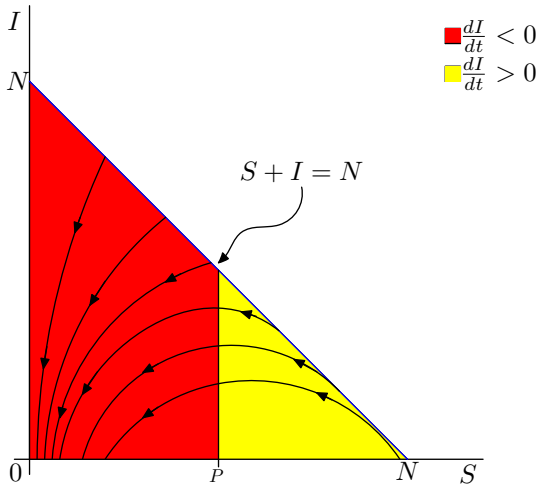

Figure 1.3: Phase Plane diagram showing the relationship between Susceptible and Infective classes in the SIR model.

Using the relationships derived, and the initial conditions shown in equation (1.5), then the phase plane diagram shown in Figure 1.3 can be constructed. To establish the severity of an epidemic

we need to calculate the maximum number of invectives, $I_{\max}$. We can see from Figure 1.3 and equation (1.7) that I achieves its maximum value at $S = p$ where $\frac{dI}{dt} = 0$. In view of (1.11),

$$I_{\max} = -p + p\ln(p) + c,$$

thus, using (1.12) we can derive,

$$\begin{aligned} I_{\max} &= -p + p\ln(p) + I_0 + S_0 - p\ln(S_0) \\ &= I_0 + S_0 - p + p\ln\left(\frac{p}{s_0}\right) \\ &= N - p + p\ln\left(\frac{p}{s_0}\right). \end{aligned}$$

Hence, $I \to 0$ as $t \to \infty$. It's also clear from the phase plane diagram shown in Figure 1.3 that $\frac{dS}{dt} < 0$ for $S \neq 0$ and $I \neq 0$. Next, by eliminating time t from (1.1) and (1.3), we get

$$\frac{dS}{dR} = -\frac{r}{a}S = -\frac{S}{p}. \tag{1.13}$$

Separate the variables and integrate both sides to get

$$\ln(S) = -\frac{R}{p} + c,$$

from which we can conclude that

$$S = Ce^{-\frac{R}{p}}, \tag{1.14}$$

with C being any real constant. By using the initial condition (1.5) we can deduce that,

$$C = S_0.$$

We can thus conclude that the solution of (1.13) is given by

$$S = S_0 e^{-\frac{R}{p}}; \tag{1.15}$$

hence, we can now use the following relation

$$S = S_0 e^{-\frac{R}{p}} \geq S_0 e^{-\frac{N}{p}} > 0,$$

to show that

$$0 < S(\infty)$$

and to conclude that

$$0 < S(\infty) \leq N.$$

Moreover from Figure 1.3 we see that the inequality is actually $0 < S(\infty) \leq p$ and $I(\infty) = 0$ implies that,

$$R(\infty) = N - S(\infty)$$

hence,

$$S(\infty) = S_0 e^{-\frac{R(\infty)}{p}} = S_0 e^{-\frac{N-S(\infty)}{p}}.$$

Therefore, as $0 < S(\infty) < p$, $S(\infty)$ is the positive root of the transcendental equation,

$$S_0 e^{-\frac{N-z}{p}} = Z, 0 < Z < p. \tag{1.16}$$

The total number who become infected is shown as the I_{total},

$$I_{\text{total}} = I_0 + S_0 - S(\infty). \tag{1.17}$$

It is apparent from equation (1.16) that $S(\infty)$ is a solution of Z. We can deduce that $I(t) \to 0$ where $S(t) \to S(\infty) > 0$, therefore the disease does not die out because of a lack of susceptibles, but due to a lack of infectives. To apply this model to actual epidemic situations we need to know the amount removed per unit time, dR/dt. From equation (1.3),(1.4) and (1.15) we can get an equation with only one parameter R,

$$\frac{dR}{dt} = aI = a(N - R - S) = a(N - R - S_0 e^{-\frac{R}{p}}), \quad R(0) = 0. \tag{1.18}$$

Kermack and McKendrick(1927) proposed using Taylor expansion around zero to evaluate the $e^{-\frac{R}{p}}$ term [9]. Using this formula we can evaluate $e^{-\frac{R}{p}}$ around the point $R = 0$,

$$\begin{aligned} e^{-\frac{R}{p}} &= 1 - \frac{1}{p}R + \frac{1}{2p^2}R^2 - \frac{1}{6p^3}R^3 + \dots \\ &\approx 1 - \frac{1}{p}R + \frac{1}{2p^2}R^2. \end{aligned}$$

Therefore,

$$-S_0 e^{-\frac{R}{p}} \approx -S_0 + \frac{S_0}{p}R - \frac{S_0}{2p^2}R^2.$$

Thus, we can re-write equation (1.18) as

$$\frac{dR}{dt} = a\left[N - S_0 + \left(\frac{S_0}{p} - 1\right)R - \frac{S_0 R^2}{2p^2}\right]. \tag{1.19}$$

Factorising the right hand side of equation (1.19) will enable us to integrate this equation in order to deduce the solution. We will now find the roots of equation (1.19) as follows,

$$\frac{dR}{dt} = a\left[N - S_0 + \left(\frac{S_0}{p} - 1\right)R - \frac{S_0 R^2}{2p^2}\right] \tag{1.20}$$

Since $\left(\frac{S_0}{p} - 1\right)^2 + 4(N - S_0)\left(\frac{S_0}{2p^2}\right) > 0$, the quadratic equation on the right hand side of (1.20) has two real roots given by

$$\begin{aligned} R_{1,2} &= \frac{-\left(\frac{S_0}{p} - 1\right) \pm \sqrt{\left(\frac{S_0}{p} - 1\right)^2 + 4(N - S_0)\left(\frac{S_0}{2p^2}\right)}}{\frac{-S_0}{p^2}} \\ &= \frac{-\left(\frac{S_0}{p} - 1\right) \pm \alpha}{\frac{-S_0}{p^2}}, \end{aligned} \tag{1.21}$$

with

$$\alpha = \sqrt{\left(\frac{S_0}{p} - 1\right)^2 + 2(N - S_0)\left(\frac{S_0}{p^2}\right)}.$$

Therefore we have the following,

$$R_1 = p - \frac{p^2}{S_0} - \frac{p^2\alpha}{S_0}, \quad \text{and} \quad R_2 = p - \frac{p^2}{S_0} + \frac{p^2\alpha}{S_0}.$$

Since $\alpha > \frac{S_0}{p} - 1$, (1.21) implies that $R_1 < 0$ and $R_2 > 0$. Thus,

$$\begin{aligned}\frac{dR}{dt} &= a\left[N - S_0 + \left(\frac{S_0}{p} - 1\right)R - \frac{S_0 R^2}{2p^2}\right] \\ &= a\frac{-S_0}{2p^2}(R - R_1)(R - R_2).\end{aligned}$$

We can solve the last equation by applying the separation of variables technique, namely

$$\int \frac{dR}{(R - R_1)(R - R_2)} = \int \frac{-S_0 a}{2p^2} dt. \tag{1.22}$$

We can easily see that,

$$\frac{2\alpha p^2}{S_0}\frac{dR}{(R - R_1)(R - R_2)} = \left[\frac{1}{R - R_2} - \frac{1}{R - R_1}\right] dR.$$

Therefore,

$$\begin{aligned}\int \frac{1}{R - R_2} - \int \frac{1}{R - R_1} &= \ln|R - R_2| - \ln|R - R_1| \\ &= \ln\left|\frac{R - R_2}{R - R_1}\right|.\end{aligned}$$

Which implies,

$$\begin{aligned}\ln\left|\frac{R - R_2}{R - R_1}\right| &= -\alpha a \int dt \\ &= -\alpha a t - \alpha a c\end{aligned}$$

where c is any constant. Using the initial conditions $R(0) = 0$ we get

$$\ln\left|\frac{-R_2}{-R_1}\right| = -\alpha a c.$$

Note that $R_1 < 0$ and $R_2 > 0$, therefore we obtain

$$R_2 = -R_1 e^{-\alpha a c}.$$

Substituting the values for R_1 and R_2 into the above equation,

$$p - \frac{p^2}{S_0} + \frac{\alpha p^2}{S_0} = -\left(p - \frac{p^2}{S_0} - \frac{\alpha p^2}{S_0}\right) e^{-\alpha a c},$$

Collecting like terms will enable us to simplify further,

$$p\left(1 + e^{-\alpha a c}\right) - \frac{p^2}{S_0}\left(1 + e^{-\alpha a c}\right) + \frac{\alpha p^2}{S_0}\left(1 - e^{-\alpha a c}\right) = 0. \tag{1.23}$$

Dividing (1.23) by $(1 + e^{-\alpha a c})$ to get that

$$p\left(1 - \frac{p}{S_0}\right)\left[\frac{1 + e^{-\alpha a c}}{1 - e^{-\alpha a c}}\right] = -\frac{\alpha p^2}{S_0}.$$

Using trigonometric identities enables us to replace the exponential function,

$$p\left(1 - \frac{p}{S_0}\right)\left[\tanh\left(\frac{-\alpha a c}{2}\right)\right]^{-1} = \frac{\alpha p^2}{S_0},$$

which implies

$$\left[\tanh\left(\frac{-\alpha a c}{2}\right)\right]^{-1} = \frac{\alpha p}{S_0\left(1 - \frac{p}{S_0}\right)}.$$

Therefore,

$$\tanh\left(\frac{-\alpha a c}{2}\right) = \left(\frac{S_0}{p} - 1\right)\frac{1}{\alpha},$$

Which implies the following

$$-\alpha a c = \tanh^{-1}\left(\frac{\left(\frac{S_0}{p} - 1\right)}{\alpha}\right) = 2\phi.$$

Now that we have calculated the constant, we can evaluate an expression for $R(t)$ which only depends on t,

$$\ln\left|\frac{R - R_2}{R - R_1}\right| = -\alpha a t + 2\phi$$

which implies

$$\ln\left|\frac{R - R_1}{R - R_2}\right| = \alpha a t - 2\phi.$$

Thus,

$$\left|\frac{R - R_1}{R - R_2}\right| = e^{\alpha a t - 2\phi}.$$

By calculating the second derivative we can determine the type of equilibrium point R_1 and R_2 are,

$$\frac{dR^2}{dt^2} = a\left[\frac{S_0}{p} - 1 - \frac{S_0 R}{p^2}\right].$$

At R_1 we have,

$$\frac{dR^2}{dt^2} = a\left[\frac{S_0}{p} - 1 - \frac{S_0}{p^2}\left(p - \frac{p^2}{S_0} - \frac{p^2\alpha}{S_0}\right)\right] = a\alpha > 0.$$

As $a, \alpha > 0$, R_1 is a minimum point. For R_2 we get

$$\frac{dR^2}{dt^2} = a\left[\frac{S_0}{p} - 1 - \frac{S_0}{p^2}\left(p - \frac{p^2}{S_0} + \frac{p^2\alpha}{S_0}\right)\right] = -a\alpha < 0,$$

which implies R_2 is a maximum point. Thus, $(R - R_1) > 0$ since $R_1 < R$ with $R_1 < 0$. Whereas $(R - R_2) < 0$ as $R < R_2$ with $R_2 > 0$. Therefore we get that

$$R - R_1 = (R - R_2)\left(-e^{\alpha a t - 2\phi}\right). \tag{1.24}$$

Substituting the values for R_1 and R_2 into (1.24) will enable us to derive an expression for $R(t)$,

$$R - \left(p - \frac{p^2}{S_0} + \frac{\alpha p^2}{S_0}\right) = \left(R - \left(p - \frac{p^2}{S_0} - \frac{\alpha p^2}{S_0}\right)\right)\left(-e^{\alpha a t - 2\phi}\right),$$

collecting like terms together will further enable us to simplify the expression,

$$R\left(1 + e^{\alpha a t - 2\phi}\right) - p\left(1 + e^{\alpha a t - 2\phi}\right) + \frac{p^2}{s_0}\left(1 + e^{\alpha a t - 2\phi}\right) - \frac{\alpha p^2}{S_0}\left(1 - e^{\alpha a t - 2\phi}\right) = 0,$$

hence,

$$\begin{aligned} R(t) &= p\left(1 - \frac{p}{S_0}\right) + \frac{\alpha p^2}{S_0}\left(\frac{1 - e^{\alpha a t - 2\phi}}{1 + e^{\alpha a t - 2\phi}}\right) \\ &= \frac{p^2}{S_0}\left[\left(\frac{S_0}{p} - 1\right) + \alpha \tanh\left(\frac{\alpha a t}{2} - \phi\right)\right]. \end{aligned} \tag{1.25}$$

Therefore, the removal rate is given by,

$$\frac{dR}{dt} = \frac{a\alpha^2 p^2}{2S_0}\operatorname{sech}^2\left(\frac{\alpha a t}{2} - \phi\right). \tag{1.26}$$

The removal rate formula (1.26) has only three parameters, namely $\frac{a\alpha^2 p^2}{2S_0}$, αa and ϕ. This simplifies the application of the model to real life data, and can means the removal rate can be quickly approximated.

3 The Bombay plague Epidemic

Kermack and McKendrick used parameters values which accurately represented the data collected in 1906, this enables the model to be a good fit when compared with the real figures [9]. Hence they came up with the following adapted removal rate.

$$\frac{dR}{dt} = 890\operatorname{sech}^2(0.2t - 3.4) \tag{1.27}$$

Since most people who got the plague died, Figure 1.4 is the approximate number of deaths per week. It shows that we would expect a peak at around 17 weeks, where 900 people would die in a single week. In comparison to the overall population size, Kermack and McKendrick concluded the epidemic was not too severe [9].

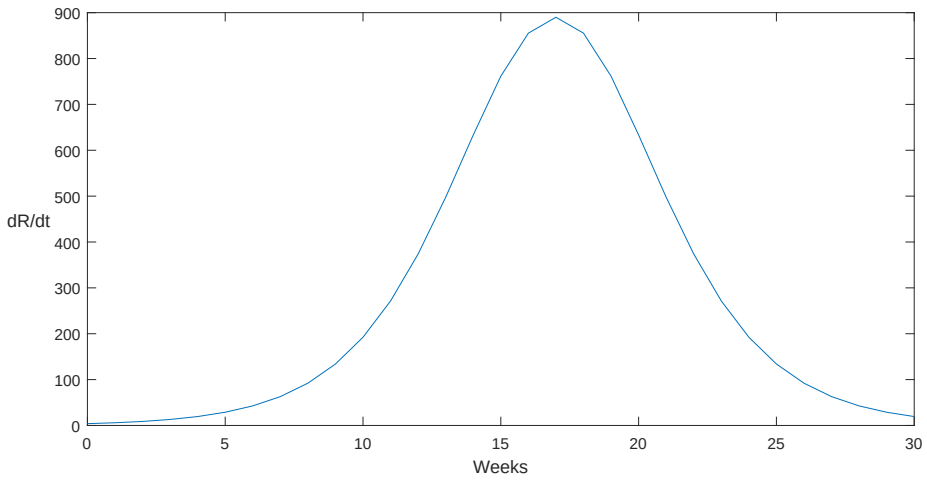

Figure 1.4: Graph of equation (1.27) over a 30 week time period.

During the Bombay plague epidemic in the spring of 1906, the incubation period was around 10 days. Death from the plague would occur around 4 days after the incubation period. Hence we can approximate that it took about 2 weeks for people to die from the plague after being infected. Hence the removal rate of invectives is 0.5 per week. Therefore if $a = 0.5$ it is possible to work out the other parameters. Using the values in equation (1.27) and comparing them to the parameters in equation (1.26) we deduce that $\alpha = 0.8$, hence

$$\tanh^{-1}\left(\frac{\frac{S_0}{p} - 1}{\alpha}\right) = 3.4$$

therefore,

$$\frac{\frac{S_0}{p} - 1}{\alpha} = \tanh(3.4)$$

which implies that,

$$S_0 = 1.7982199p. \tag{1.28}$$

Comparing the other values gives,

$$\frac{a\alpha^2 p^2}{2S_0} = 890.$$

We can simplify this equation by substituting in our value for α, hence

$$\frac{p^2}{S_0} = 10002.59819,$$

using (1.28) we can get a value for p of 10002.59843. Therefore $S_0 = 17986.87154$. These values will be extremely important when we construct a numerical approximation of the model.

Chapter 2

Mathematical Modeling of HIV

1 Introduction to Human Immunodeficiency Virus (HIV)

Sexually transmitted diseases (STDs), such as Chlamydia, Gonorrhea and HIV, differ from many other infectious diseases due to the method of transmission. The most common mode of transmission is through sexual contact. Public Health England state that 95% of those who where infected with HIV in England during 2013 acquired the disease through sexual contact.

Before we discuss the mathematical modelling of STDs such as HIV, it is first necessary to deduce the factors that affect the spread of the disease and the ways in which it is transmitted. Firstly we need to consider the interactions between different susceptibles. In a completely heterosexual population, it is a criss cross disease as there are two interacting groups, males and females, where each group is the disease host for the other. However, in real populations there exists homosexual, bisexual and heterosexual populations.

In 2013 the British Office of National statistics (ONS) conducted the integrated household survey and found the following statistics [13]:

- 1.2 per cent of adults identified themselves as gay or lesbian which is approximately 545,000 adults.
- 0.5 per cent of adults identified themselves as bisexual which equates to approximately 220,000 adults.
- 3.9 per cent of adults said they they didn't know or refused to give an answer on there sexuality.

Of the statistics gathered by the ONS, one of the most surprising is that men were more likely than women to consider themselves as gay or lesbian. Indeed, 1.6 per cent of adult males identified themselves as gay compared with just 0.8 per cent of adult females.

When HIV antibodies are detected in the patient, the patient is categorised as HIV positive, these antibodies can be produced anytime from a week to several months after the infection. Unless there is an intervention with drug therapy the patient will enter a latent period before exhibiting the

end-stage disease which is classified as acquired immunodeficiency syndrome (AIDS). According to the World Health Organization it can take up to 10 years for HIV to become AIDS, however in the majority of cases it takes a significantly shorter period of time. This varying incubation period is a major issue when it comes to understanding and controlling the spread of the disease.

The rise in the quantity of people who have STD's is a major worry for not only the world health organisations but also for governments around the world. Certain characteristics of STD's such as only the sexually active population can be infected may appear to make any epidemic outbreaks easier to control. However many social problems such as the stigma attached to HIV means that the virus is not openly discussed within society and hence gathering any accurate data is extremely difficult. A report written by UNAIDS on the global epidemic of HIV and AIDS estimates that 27 million people are HIV positive but are unaware that they contain the virus.[20]

Data which is available shows that the majority of people who already have HIV and AIDS live in some of the poorest countries in the world. Recent estimates suggest that as many as 70% of deaths due to AIDS occur in Africa. The report by UNAIDS [20] also highlights how the inability to control the spread of AIDS is an index of the substandard governance, bad education and a reluctance to accelerate the development of underdeveloped countries. The decline of real per capita incomes in the 30 poorest economies since 1980 is just one example of the failure of the world organisations to help improve the lives of millions of people. Unless developed countries open their markets to less developed countries and begin to try and encourage globalisation, it is impossible for the poorest countries to make significant strides in tackling poverty. Reports such as the one by UNAIDs have highlighted that it is not feasible to try and just tackle each disease independently. To be successful it is necessary to tackle the more general issues such as poverty, income inequality and refugee movement, all of which affect the spread of HIV. Tackling the wider issues will mean more developed countries will posses a greater ability to not only tackle the spread of STD's such HIV, but to educate more people on how to avoid catching such diseases, and where to go if they think they have the symptoms.

If we imagine a population in which everyone is infected with HIV at $t = 0$, and let $z(t)$ denote the proportion of people who have AIDS, $v(t)$ as the proportion who are HIV positive but do not yet have AIDS and k(t) be the rate of conversion from HIV positive to aids, then we have the following model,

$$\begin{aligned} \frac{dv}{dt} &= -k(t)v, \\ \frac{dz}{dt} &= k(t)v, \end{aligned} \tag{2.1}$$

where

$$v(0) = 1, \quad z(0) = 0, \quad v + z = 1.$$

This simplified model makes the assumption that every person who is HIV positive develops AIDS, which is not always the case. From biological research it is known that the immune system's response to diseases becomes increasingly impaired the longer the period of time since infection. Hence the

conversion rate to AIDS is an increasing function,

$$k(t) = bt,$$

where $b > 0$ is constant. Thus, the first equation of (2.1) can be written in the form

$$\frac{dv}{dt} = -btv, \tag{2.2}$$

which implies

$$\ln(v) = \frac{-bt^2}{2} + c_1,$$

with c_1 being any real constant. We can easily see that

$$v(t) = C_1 e^{\left(\frac{-bt^2}{2}\right)},$$

by using the initial condition $v(0) = 1$, we get that

$$v(t) = e^{\left(\frac{-bt^2}{2}\right)}.$$

Since $v + z = 1$, we can deduce that

$$z(t) = 1 - e^{\left(\frac{-bt^2}{2}\right)}.$$

However this model is very simple and has unrealistic assumptions such as everyone is infected with HIV at $t = 0$. Therefore in this next section we will use a more realistic model.

2 Modeling an Aids Epidemic in a Homosexual Population

If we now look at the development of an aids epidemic in a homosexual population where $X(t)$, $Y(t)$, $A(t)$ and $Z(t)$ denote the amount of susceptible's, infectious males, AIDS patients and the amount of seropositives who are non-infectious, we can make a more realistic model by making the following realistic assumptions,

- We let B represent the birth rate entering the susceptible class.
- We assume that individuals in all classes can die naturally and we denote this natural death rate as μ.
- We also let the rate of transmission from susceptible to infectious be λc, where λ represents the probability of being infected with HIV from a random individual, and c represents the number of sexual partners an individual has had.
- We denote the proportion of infectives who develop AIDS as pv, where p is the proportion of HIV positive individuals who are infectious.
- Hence, the proportion of infectives who don't develop the disease and become non-infectious is $(1 - p)v$.

- Finally we let the death rate due to aids be denoted by d.

These assumptions enable us to create Figure 2.1 representing the flow of individuals depending on their infectious status.

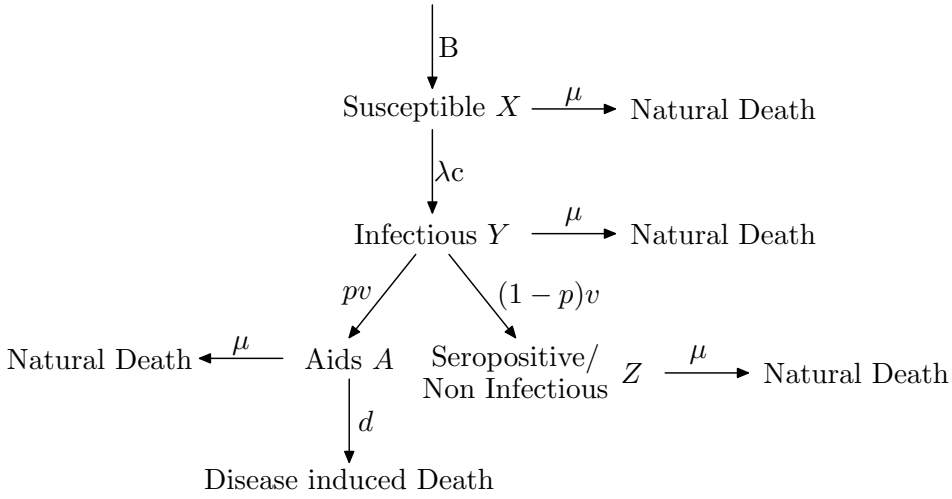

Figure 2.1: Flow chart showing the different classes and the parameters which effect the movement of individuals from one class to another.

We can thus arrive at the following system of ordinary differential equations which models HIV disease

$$\frac{dX}{dt} = B - \mu X - \lambda c X, \quad \lambda = \frac{\beta Y}{N} \tag{2.3}$$

$$\frac{dY}{dt} = \lambda c X - (v + \mu) Y, \tag{2.4}$$

$$\frac{dA}{dt} = pvY - (d + \mu) A, \tag{2.5}$$

$$\frac{dZ}{dt} = (1 - p) vY - \mu Z, \tag{2.6}$$

see [12] for more details. Here, β is the transmission probability and the total population size can be expressed as,

$$N(t) = X(t) + Y(t) + Z(t) + A(t). \tag{2.7}$$

From these parameters shown in the model above, we are able to deduce that the average incubation time of the disease is $1/v$. Thus, the basic reproduction ratio can be defined as,

$$R_0 \approx \frac{\beta c}{v}. \tag{2.8}$$

The epidemic begins when $R_0 > 1$, where the population model will evolve to a steady state as shown below. Steady states occur when,

$$B - \mu X - \lambda c X = 0, \tag{2.9}$$

$$\lambda cX - (v+\mu)Y = 0, \tag{2.10}$$
$$pvY - (d+\mu)A = 0, \tag{2.11}$$
$$(1-p)vY - \mu Z = 0. \tag{2.12}$$

By adding (2.9) and (2.10) we obtain

$$B - \mu X - (v+\mu)Y = 0,$$

from which we can conclude that

$$Y = \frac{B-\mu X}{v+\mu}. \tag{2.13}$$

The re-arranging of (2.9) produces a second equation for Y, that is

$$\begin{aligned} B &= X(\mu + \lambda c) \\ &= X(\mu + \frac{\beta c Y}{N}), \end{aligned}$$

which implies

$$Y = \frac{NB}{\beta cX} - \frac{\mu N}{\beta c}. \tag{2.14}$$

In view of (2.13) and (2.14) we have that

$$\frac{NB}{\beta cX} - \frac{\mu N}{\beta c} = \frac{B-\mu X}{v+\mu},$$

from which we get

$$BX - \mu X^2 + \frac{(v+\mu)\mu NX}{\beta c} = \frac{(v+\mu)NB}{\beta c},$$

i.e.,

$$X^2 - X\left(\frac{(v+\mu)\mu N}{\beta c} + \frac{B}{\mu}\right) + \frac{(v+\mu)NB}{\beta\mu c} = 0. \tag{2.15}$$

Therefore

$$X = \frac{\left(\frac{(v+\mu)N}{\beta c} + \frac{B}{\mu}\right) \pm \sqrt{\left(\frac{(v+\mu)N}{\beta c} + \frac{B}{\mu}\right)^2 - 4\frac{(v+\mu)NB}{\beta\mu c}}}{2} \tag{2.16}$$

The value inside the square root in (2.16) can be simplified as follows

$$\begin{aligned} \left(\frac{(v+\mu)N}{\beta c} + \frac{B}{\mu}\right)^2 - 4\frac{(v+\mu)NB}{\beta\mu c} &= \left(\frac{N^2(v+\mu)^2}{\beta^2c^2} + \frac{B^2}{\mu^2} + 2\frac{NB(v+\mu)}{\beta\mu c}\right) - 4\frac{(v+\mu)NB}{\beta\mu c} \\ &= \left(\frac{N^2(v+\mu)^2}{\beta^2c^2} + \frac{B^2}{\mu^2} - 2\frac{NB(v+\mu)}{\beta\mu c}\right) \\ &= \left(\frac{N(v+\mu)}{\beta c} - \frac{B}{\mu}\right)^2. \end{aligned} \tag{2.17}$$

From which we can conclude that (2.16) has two real roots given by

$$X = \frac{\left(\frac{(v+\mu)N}{\beta c} + \frac{B}{\mu}\right) \pm \left|\frac{N(v+\mu)}{\beta c} - \frac{B}{\mu}\right|}{2}.$$

We can easily see that for any real numbers x and y there holds $x + y \pm |x - y| = 2x$ or $2y$. We can thus deduce that the last equation has the following two solutions

$$X^* = \frac{B}{\mu} \quad \text{and} \quad X^* = \frac{(v+\mu)N}{\beta c}. \tag{2.18}$$

Note that the first solution $X = \frac{B}{\mu}$, corresponds to the steady state in the absence of infection. We are interested in what happens when the disease is present, therefore we will concentrate on the second solution for X. Adding equations (2.9)-(2.12) and using (2.7) to get,

$$B - \mu N - dA = 0,$$

namely

$$A^* = \frac{B - \mu N}{d}. \tag{2.19}$$

According to (2.11) we have

$$Y = \frac{(d+\mu)A}{pv},$$

which in view of (2.19) gives

$$Y^* = \frac{(d+\mu)(B - \mu N)}{pvd}. \tag{2.20}$$

We can now use (2.12) to get

$$Z = \frac{(1-p)vY}{\mu}, \tag{2.21}$$

from which and (2.20) we can conclude that

$$Z^* = \frac{(1-p)(d+\mu)(B - \mu N)}{pd\mu}. \tag{2.22}$$

In view of (2.7), (2.18), (2.19), (2.20) and (2.22) we obtain

$$N^* = \frac{B\beta[\mu(v+d+\mu) + vd(1-p)]}{[v+\mu][\beta(d+\mu) - pv]}$$

Hence, the infected steady state occurs at,

$$\begin{aligned} X^* &= \frac{(v+\mu)N}{\beta c}, \quad Y^* = \frac{(d+\mu)(B - \mu N^*)}{pvd}, \\ Z^* &= \frac{(1-p)(d+\mu)(B - \mu N^*)}{pd\mu}, \quad A^* = \frac{B - \mu N^*}{d}, \\ N^* &= \frac{B\beta[\mu(v+d+\mu) + vd(1-p)]}{[v+\mu][\beta(d+\mu) - pv]}. \end{aligned} \tag{2.23}$$

To calculate the stability of the infected steady states we first need to calculate the Jacobian J matrix of the system of equations (2.3)-(2.6) at the critical point (X^*, Y^*, A^*, Z^*). We have that

$$J(X^*, Y^*, A^*, Z^*) = \begin{pmatrix} -\mu - \lambda c & \frac{-c\beta X^*}{N^*} & 0 & 0 \\ \lambda c & \frac{c\beta X^*}{N^*} - (v+\mu) & 0 & 0 \\ 0 & pv & -(d+\mu) & 0 \\ 0 & (1-p)v & 0 & -\mu \end{pmatrix}.$$

Next we evaluate the eigenvalues δ, of matrix J, which satisfy the equation

$$|J - \delta I| = 0 \quad \Rightarrow 0 = \begin{vmatrix} (-\mu - \lambda c) - \delta & \frac{-c\beta X^*}{N^*} & 0 & 0 \\ \lambda c & \frac{c\beta X^*}{N^*} - (v+\mu) - \delta & 0 & 0 \\ 0 & pv & -(d+\mu) - \delta & 0 \\ 0 & (1-p)v & 0 & -\mu - \delta \end{vmatrix}.$$

We can deduce that the first eigenvalue is $\delta_1 = -\mu$ and the other eigenvalues must satisfy

$$0 = \begin{vmatrix} (-\mu - \lambda c) - \delta & \frac{-c\beta X^*}{N^*} & 0 \\ \lambda c & \frac{c\beta X^*}{N^*} - (v+\mu) - \delta & 0 \\ 0 & pv & -(d+\mu) - \delta \end{vmatrix}, \tag{2.24}$$

from which we can conclude that

$$(-(d+\mu) - \delta)\left[((-\mu - \lambda c) - \delta)\left(\frac{c\beta X^*}{N^*} - (v+\mu) - \delta\right) - (\lambda c)\left(\frac{-c\beta X^*}{N^*}\right)\right] = 0.$$

Hence, we can deduce that the second eigenvalue is $\delta_2 = -(d+\mu)$. Moreover,

$$\begin{aligned} 0 &= ((-\mu - \lambda c) - \delta)\left(\frac{c\beta X^*}{N^*} - (v+\mu) - \delta\right) - (\lambda c)\left(\frac{-c\beta X^*}{N^*}\right) \\ &= \delta^2 + \left((v+\mu) + (\mu + \lambda c) - \frac{c\beta X^*}{N^*}\right)\delta \\ &\quad + \left[(-\mu - \lambda c)\left(\frac{c\beta X^*}{N^*} - (v+\mu)\right) + (\lambda c)\left(\frac{c\beta X^*}{N^*}\right)\right] \end{aligned} \tag{2.25}$$

According to (2.18) the constant term in the above equation can be simplified as follows

$$\begin{aligned} (-\mu - \lambda c)\left(\frac{c\beta X^*}{N^*} - (v+\mu)\right) + (\lambda c)\left(\frac{c\beta X^*}{N^*}\right) &= -(\mu + \lambda c)(v+\mu) + (v+\mu)(\mu + \lambda c) + \lambda c(v+\mu) \\ &= \lambda c(v+\mu) \end{aligned}$$

Hence, (2.25) becomes,

$$\delta^2 + ((v+\mu) + (\mu + \lambda c) - (v+\mu))\delta + \lambda c(v+\mu) = 0$$

i.e,

$$\delta^2 + (\mu + \lambda c)\delta + \lambda c(v+\mu) = 0.$$

We can apply the quadratic formula to find the third and fourth eigenvalues, namely

$$\delta_{3,4} = \frac{-(\mu+\lambda c) \pm \sqrt{(\mu+\lambda c)^2 - 4\lambda c(v+\mu)}}{2}.$$

To evaluate the stability of the infected steady state we do not require the exact solution for each eigenvalue, instead we just need to obtain whether or not they are positive or negative. From inspection it is apparent that δ_1 and δ_2 are both negative since the parameters d and μ are both positive. If δ_3 and δ_4 are both negative then the steady state is stable.If $\delta_3 > 0$ then

$$\delta_3 = \frac{-(\mu+\lambda c) + \sqrt{(\mu+\lambda c)^2 - 4\lambda c(v+\mu)}}{2} > 0, \tag{2.26}$$

which implies that

$$(\mu+\lambda c) < \sqrt{(\mu+\lambda c)^2 - 4\lambda c(v+\mu)}.$$

Moreover,

$$(\mu+\lambda c)^2 - (\mu+\lambda c)^2 < -4\lambda c(v+\mu),$$

which gives,

$$0 < -\lambda c(v+\mu). \tag{2.27}$$

The inequality (2.27) is not true since all four parameters λ,c,v and μ are positive, therefore $\delta_3 < 0$. Similarly for δ_4 to be positive,

$$\delta_4 = \frac{-(\mu+\lambda c) - \sqrt{(\mu+\lambda c)^2 - 4\lambda c(v+\mu)}}{2} > 0,$$

then

$$(\mu+\lambda c) < -\sqrt{(\mu+\lambda c)^2 - 4\lambda c(v+\mu)},$$

i.e.,

$$(\mu+\lambda c)^2 - (\mu+\lambda c)^2 < -4\lambda c(v+\mu),$$

from which we can conclude that

$$0 < -\lambda c(v+\mu).$$

This is the same condition, and cannot be true since all the parameters involved are positive. Therefore δ_4 is also negative, thus the infected steady state is stable. To deduce the type of the infected steady state, we need to determine whether or not the eigenvalues are real or imaginary. Therefore we will evaluate the sign of the square root in (2.26), If steady state is complex then we get that,

$$(\mu+\lambda c)^2 - 4\lambda c(v+\mu) < 0. \tag{2.28}$$

If (2.28) is true then the infected steady state is complex, however it will still be stable as the real part of the eigenvalues is still negative. Earlier we deduced a value for R_0 in equation (2.8). During the early stages in an epidemic, almost every individual is in the Susceptible class, where $X \approx N$, hence

$$\frac{dY}{dt} \approx (\beta c - v - \mu)Y$$

where

$$R_0 \approx \frac{\beta c}{v} \Rightarrow \frac{dY}{dt} \approx v(R_0 - 1)Y \tag{2.29}$$

This relation exists since the average life expectancy of a susceptible individual is a lot longer than the average incubation time from the point of infection to the acquisition of the disease, therefore $v \gg \mu$. We will now solve equation (2.29) to obtain a solution for $Y(t)$,

$$\frac{dY}{dt} = v(R_0 - 1)Y.$$

Separating the variables will enable us to integrate both sides of the equation

$$\Rightarrow \int \frac{1}{Y} dY = \int v(R_0 - 1)dt$$

from which we can conclude that

$$\Rightarrow \ln(Y) = v(R_0 - 1)t + c_2$$

where c_2 is a constant. Therefore,

$$Y(t) = Y(0)e^{rt} \tag{2.30}$$

where $r = v(R_0 - 1)$. It is now possible to calculate the doubling time t_d for the infectious males class, when an epidemic exists the doubling time occurs when $y(t_d) = 2Y(0)$. Therefore we can see that from equation (2.30),

$$Y(t_d) = Y(0)e^{rt_d}$$

hence

$$2Y(0) = Y(0)e^{rt_d}$$

from which we can conclude the doubling time is

$$t_d = \frac{\ln(2)}{v(R_0 - 1)} \tag{2.31}$$

From equation (2.31) we can see that as R_0 increases the doubling time of the epidemic shortens. Substituting our solution for Y(t) into our equation for the population of AIDS patients (2.5), we get,

$$\frac{dA}{dt} = pvY(0)e^{rt} - (d + \mu)A \tag{2.32}$$

using an integrating factor we can evaluate the solution of equation (2.32),

$$\frac{dA}{dt} + (d + \mu)A = pvY(0)e^{rt}$$

if we implement an integrating factor $s(t)$, where $s(t) = e^{\int (d+\mu)dt} = e^{(d+\mu)t}$,then

$$e^{(d+\mu)t}\frac{dA}{dt} + e^{(d+\mu)t}(d + \mu)A = pvY(0)e^{rt}e^{(d+\mu)t}$$

which implies that

$$\frac{d}{dt}(e^{(d+\mu)t}A) = pvY(0)e^{rt+(d+\mu)t}$$

we can now integrate both sides

$$\int \frac{d}{dt}(e^{(d+\mu)t}A)dt = \int pvY(0)e^{(r+d+\mu)t}dt$$

which gives

$$e^{(d+\mu)t}A = pvY(0)\frac{e^{(r+d+\mu)t}}{r+d+\mu} + c_3$$

where c_3 is any real constant. Therefore

$$A = pvY(0)\frac{e^{rt}}{r+d+\mu} + c_3e^{-(d+\mu)t}$$

At the beginning of the epidemic there exists no aids patients as there is a reasonable incubation period before HIV patients can develop aids, which implies $A(0) = 0$, hence

$$\begin{aligned} 0 &= pvY(0)\frac{e^{r(0)}}{r+d+\mu} + c_3e^{-(d+\mu)(0)} \\ &= pvY(0)\frac{1}{r+d+\mu} + c_3 \end{aligned}$$

we can now evaluate c_3,

$$c_3 = -pvY(0)\frac{1}{r+d+\mu}$$

therefore we can write $A(t)$ as,

$$\begin{aligned} A(t) &= pvY(0)\frac{e^{rt}}{r+d+\mu} - pvY(0)\frac{1}{r+d+\mu}e^{-(d+\mu)t} \\ &= pvY(0)\frac{e^{rt} - e^{-(d+\mu)t}}{r+d+\mu}. \end{aligned} \tag{2.33}$$

Anderson [2] wrote a paper on transmission dynamics of HIV and used data collected by Peterman [15] which focused on the spread and distribution of HIV in several different populations. The report written by Anderson highlighted the key figures from the data collected by Peterman such as the rate r, and the doubling time t_d. Some of the figures are shown in the table 2.1.

Country	Period	Rate (r/yr)	Doubling Time t_d
USA	1981-1985	0.9	9.2
England	1982-1985	1.27	6.6
Italy	1983-1985	1.66	5
Switzerland	1983-1985	0.84	9.9
Sweden	1983-1985	1.04	8

Table 2.1: Doubling time t_d from HIV in the early stages of an epidemic

If we look at the data above it is clear that the doubling time varies quite a lot depending on which country we are investigating. Using the data from 6875 homosexual and bisexual men who had

attended a clinic over a 5 year period from 1978 to 1985 in America, we can calculate accurate values for the parameters in equation (2.33). They calculated $r = 0.88yr^{-1}$,$R_0 \approx 5.15$,$d + \mu \approx d = 1 - 1.33yr^{-1}$,$p = 10\% - 30\%$,$v \approx 0.22yr^{-1}$ and $c = 2 - 6$ partners a month. If we substitute these into equation (2.31) we can estimate the doubling time for the HIV positive class of individuals,

$$\begin{aligned} t_d &= \frac{\ln(2)}{0.22(5.15-1)} \\ t_d &= 0.7592\text{years} \end{aligned} \tag{2.34}$$

The doubling time for the aids class is approximately 9 months. This figure is similar to the data for the USA as whole shown in Table 2.1. We can see from the calculations that the doubling time in dependent onR_0, The larger R_0 is the shorter the doubling time for the HIV positive class. This relationship is to be expected if the epidemic is very severe, as there will be a greater number of secondary infections from an average primary case due to an increase in the transmission probability or the contact rate. Hence doubling time will be a lot shorter. If we analyse some of the data displayed in the report by Anderson [2] in more detail we can see from table 2.2 that the doubling time of aids incidence is greater in homosexual/bisexual men than it is in Heterosexual contacts in the USA.

Group	Doubling Time t_d	Cases (1986)	% of Total
Homosexual/Bisexual men, IV Drug Abusers	9.5	599	6.9
Homosexual/Bisexual men, non-IV Drug Abusers	9.1	5009	65.4
IV Drug Abusers	8.25	1429	16.5
Heterosexual Contacts	7.85	100	1.1

Table 2.2: Doubling time t_d in aids incidence by risk group in the early stages of an epidemic

In this chapter we have established that within the HIV model, there exists an infected steady state. Therefore it is vitally important to understand the biological makeup of HIV, and to establish if there exists some treatment which will eradicate the virus. This is what we will now focus on in the next chapter.

Chapter 3

Treatment for HIV and Aids

1 Biological Makeup of HIV

Unlike many other retroviruses, HIV uses the mRNA of the invaded cell to synthesise its own RNA. There are two main known varieties of HIV, the most common of which is HIV-1. Whilst most viruses are destroyed by the immune system, HIV is only temporally halted by it and so is never destroyed. Although HIV is one of the most studied viruses, some of its characteristics are not fully understood, such as the reason for the sudden drop in T-cell count after the virus has infected the host cells. The body usually replenishes the amount of T-cells at a quicker rate than HIV can kill them. The normal level of T-cell count in a uninfected person is around $1000/\mu L$, a patient is diagnosed with aids when the level drops below $200/\mu L$. CD4-T-cells are vital in the functions of the immune system, HIV mainly infects CD4-T-cells called lymphocytes, these are a type of white blood cell."HIV attacks particular lymphocytes called T4 helper lymphocytes. The virus binds specifically to the CD4 receptors on these cells using the glycoprotein spikes on the surface of the virus particles." [6]. Although the patient immune system does produce antibodies to attack the virus, it is only successful at constraining the virus. After a medium of 8-10 years the patient develops aids. The production of antibodies to fight HIV is when the patient is categorised as HIV positive.

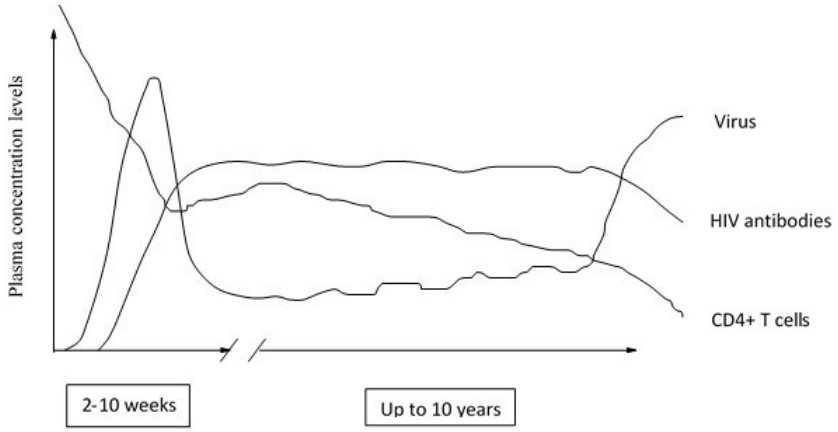

Figure 3.1: Development of HIV over short and long term time scale, alongside the change in CD4+ T cell count and the virus level. Each variable is measured as the plasma concentration level [6].

Figure 3.1 shows how the patients immune system responds to the introduction of the virus by producing HIV antibodies, this initially works as it decreases the virus plasma concentration level, however, it does not destroy the virus. The introduction of the virus coincides with a drop in the CD4+ T cell plasma concentration, after the initial reduction the rate of decrease in concentration slows. When the patient has been infected with the virus for 8-10 years, the level of CD4+ T cell plasma concentration reduces further as the virus awakes from a dormant period, this coincides with a decrease in the HIV antibody concentration level.

Whilst the virus is in a semi steady state, the patient doesn't display any symptoms of the virus. Therefore, it is important to diagnose the patient with HIV during the first 2-10 weeks. Failure to diagnose the patient will prevent any drug therapy being used to slow down the disease, whilst simultaneously increasing the likelihood of the patient infecting other individuals with HIV. Education as mentioned earlier in Chapter 2 is essential to slow the spread of a virus such as HIV.

2 Drug Therapy to Treat HIV

2.1 Linear Model to Predict the Quantity of Viral Peptides Produced

David Ho prescribed a HIV protease inhibitor to 20 subjects who had the HIV-1 virus and recorded how they responded to treatment [7]. He proposed a simple linear first order equation to model how the introduction of protease inhibitor may reduce the amount of viral peptides. He used the following

$$\frac{dV}{dt} = p - cv,$$

where p is the source of viral peptides, and c the viral clearance rate. If we assume that the drug is completely effective then $p = 0$. Therefore,

$$\frac{dV}{dt} = -cv,$$

so

$$V(t) = V(0)e^{-ct}. \tag{3.1}$$

Ho then used data collected from his subjects to calculate a half life of $t_{\frac{1}{2}} = 2.1 \pm 0.4$ days. He then substituted this into (3.1) and found that the model predicts that one billion viral particles are produced daily, contrary to previous assumptions that, when the disease was in its semi steady state, it was in-active. Hence, even a simple linear model can help establish how HIV develops over time.

2.2 Combination Drug Therapy

One of the main problems with trying to treat HIV is that it mutates to drug resistant forms and therefore it is not possible to treat patients using only one drug in the long term. The most effective treatment is to use a combination of different drugs. This means that the virus will take longer to evolve into a multi drug resistant form. Azidothymidine (AZT) is a transcriptase inhibitor that

targets the protein enzymes and makes the virus non-infectious.

The mathematical model Perelson and Nelson in their paper Mathematical Analysis of HIV-1 [6], proposed a suitable combination drug therapy model. The model proposed is a four species model that includes equations with the following unknowns

- the uninfected T-cells, T,
- the productively infected T-cells, T^*,
- the infectious viruses, V_I, and
- the noninfectious viruses, V_{NI}.

The system of differential equations describing the model is

$$\begin{aligned}
\frac{dT}{dt} &= s + pT\left(1 - \frac{T}{T_{max}}\right) - d_T T - kV_I T = f(T, T^*, V_I, V_{NI}),\\
\frac{dT^*}{dt} &= (1 - n_{rt})kV_I T - \delta T^* = g(T, T^*, V_I, V_{NI}),\\
\frac{dV_I}{dt} &= (1 - n_p)N\delta T^* - cV_I = h(T, T^*, V_I, V_{NI}),\\
\frac{dV_{NI}}{dt} &= n_p N\delta T^* - cV_{NI} = z(T, T^*, V_I, V_{NI}).
\end{aligned} \tag{3.2}$$

Here n_p and n_{rt} denote, respectively, the measure of effectiveness of the inhibitor prescribed to block the production of virus particles and of the RT-inhibitor taken. RT-inhibitor is a reverse transcriptase drug such as AZT. Where s, p, T_{max}, d_T and k are all positive constants, and s represents the source of the viral particles and $-d_T$ is the clearance term. We can see from the model that, when $n_{rt} = 1$ no T-cells are produced, therefore the drug is completely effective. Similarly, if $n_{rt} = 0$, no RT-inhibitor has been administered and T-cells will be produced.

Stability of the model We are interested in how this model behaves, analysing the steady states of the system to enables us to deduce the likely model predictions. First, we find the uninfected steady state (T_{s1},0,0,0), this occurs when

$$\begin{aligned}
0 &= s + pT\left(1 - \frac{T}{T_{max}}\right) - d_T T\\
&= -\frac{p}{T_{max}}T^2 + (p - d_T)T + s,
\end{aligned}$$

then

$$\begin{aligned}
T_{s1} &= \frac{(d_T - p) \pm \sqrt{(p - d_T)^2 + \frac{4sp}{T_{max}}}}{\frac{2p}{T_{max}}}\\
&= \frac{T_{max}}{2p}\left[(d_T - p) \pm \sqrt{(p - d_T)^2 + \frac{4sp}{T_{max}}}\right].
\end{aligned} \tag{3.3}$$

We can easily see that one of the above roots is negative, which is rejected, since we are interested only in positive steady states. Indeed,

$$\frac{T_{max}}{2p}\left[(d_T-p)-\sqrt{(p-d_T)^2+\frac{4sp}{T_{max}}}\right]>0$$

implies that

$$(d_T-p)-\sqrt{(p-d_T)^2+\frac{4sp}{T_{max}}}>0.$$

Thus,

$$(d_T-p)^2>(p-d_T)^2+\frac{4sp}{T_{max}},$$

which gives the following condition

$$\frac{4sp}{T_{max}}<0. \tag{3.4}$$

Since (3.4) cannot be true as the parameters s, p and $T_{\max}$ are all positive, there is only one positive uninfected steady state,

$$T_{s1}=\frac{T_{max}}{2p}\left[(d_T-p)+\sqrt{(p-d_T)^2+\frac{4sp}{T_{max}}}\right]. \tag{3.5}$$

We next study the stability of the above steady state. To do this, we first calculate the Jacobian matrix A, of the system of equations (3.2),

$$A=\begin{pmatrix} p\left(1-\frac{2T}{T_{max}}\right)-d_T-kV_I & 0 & -kT & 0\\ (1-n_{rt})kV_I & -\delta & (1-n_{rt})kT & 0\\ 0 & (1-n_p)N\delta & -c & 0\\ 0 & n_pN\delta & 0 & -c\end{pmatrix}.$$

Evaluating the Jacobian matrix at the steady state, we obtain

$$A(T_{s1},0,0,0)=\begin{pmatrix} p\left(1-\frac{2T_{s1}}{T_{max}}\right)-d_T & 0 & -kT_{s1} & 0\\ 0 & -\delta & (1-n_{rt})kT_{s1} & 0\\ 0 & (1-n_p)N\delta & -c & 0\\ 0 & n_pN\delta & 0 & -c\end{pmatrix}. \tag{3.6}$$

Next, we calculate the eigenvalues of matrix (3.6),

$$\begin{aligned}|A-\lambda I|=0 \quad \Rightarrow 0&=\begin{vmatrix} p\left(1-\frac{2T_{s1}}{T_{max}}\right)-d_T-\lambda & 0 & -kT_{s1} & 0\\ 0 & -\delta-\lambda & (1-n_{rt})kT_{s1} & 0\\ 0 & (1-n_p)N\delta & -c-\lambda & 0\\ 0 & n_pN\delta & 0 & -c-\lambda\end{vmatrix}\\ &=\left[p\left(1-\frac{2T_{s1}}{T_{max}}\right)-d_T-\lambda\right]\begin{vmatrix} -\delta-\lambda & (1-n_{rt})kT_{s1} & 0\\ (1-n_p)N\delta & -c-\lambda & 0\\ n_pN\delta & 0 & -c-\lambda\end{vmatrix}\\ &=\left[p\left(1-\frac{2T_{s1}}{T_{max}}\right)-d_T-\lambda\right][-(1-n_{rt})kT_{s1}[(1-n_p)N\delta(-c-\lambda)]\\ &\quad+(-c-\lambda)^2(-\delta-\lambda)].\end{aligned} \tag{3.7}$$

Hence, we can see that

$$\lambda_1 = p(1 - \frac{2T_{s1}}{T_{max}}) - d_T.$$

Moreover,

$$\begin{aligned} 0 &= (1-n_{rt})kT_{s1}[(1-n_p)N\delta(-c-\lambda)] - (-c-\lambda)^2(-\delta-\lambda) \\ &= (-c-\lambda)[(1-n_{rt})(1-n_p)kT_{s1}N\delta - (-c-\lambda)(-\delta-\lambda)] \\ &= (-c-\lambda)[(1-n_c)kT_{s1}N\delta - (-c-\lambda)(-\delta-\lambda)], \end{aligned} \tag{3.8}$$

where $n_c = (1-n_{rt})(1-n_p)$ represents the effectiveness of combination drug therapy treatment. Thus, the second eigenvalue is

$$\lambda_2 = -c.$$

Next, we calculate the other two eigenvalues. We have

$$\begin{aligned} 0 &= (1-n_c)kT_{s1}N\delta - (-c-\lambda)(-\delta-\lambda) \\ &= \lambda^2 + (c+\delta)\lambda - (1-n_c)N\delta kT_{s1} + c\delta. \end{aligned}$$

Which implies that

$$\begin{aligned} \lambda_{3,4} &= \frac{-(c+\delta) \pm \sqrt{(c+\delta)^2 + 4(1-n_c)N\delta kT_{s1} - 4c\delta}}{2} \\ &= \frac{-(c+\delta)}{2} \pm \frac{1}{2}\sqrt{(c+\delta)^2 + 4(1-n_c)N\delta kT_{s1} - 4c\delta}. \end{aligned}$$

To establish whether the steady states are imaginary or real, we analyse the sign of the square root term.

$$(c+\delta)^2 + 4(1-n_c)N\delta kT_{s1} - 4c\delta = (c-\delta^2) + 4(1-n_c)N\delta kT_{s1},$$

which is a positive quantity as $(1-n_c)$, N, δ, k and T_{s1} are all positive. We can thus conclude that the third and fourth steady states are also real. For the steady state to be stable we require the eigenvalues to be negative. From an initial inspection it is apparent that λ_2 and λ_4 are both negative. We can also see that λ_1 is negative if

$$T_{s1} > \frac{1}{2p}(p - d_T)T_{max},$$

which is satisfied from (3.5). Hence, we will next investigate the sign of the fourth eigenvalue. Let us assume that it is negative, namely

$$\frac{-(c+\delta)}{2} + \frac{1}{2}\sqrt{(c+\delta)^2 + 4(1-n_c)N\delta kT_{s1} - 4c\delta} < 0,$$

which implies

$$(c+\delta)^2 + 4(1-n_c)N\delta kT_{s1} - 4c\delta < (c+\delta)^2.$$

Thus,

$$4(1-n_c)N\delta kT_{s1} - 4c\delta < 0.$$

Re-arranging yields,

$$1 - \frac{c}{NkT_{s1}} < n_c. \tag{3.9}$$

We can thus conclude that the steady state T_{s1} is stable under the assumption (3.9). The stability condition says that if the treatment is strong enough, the virus level will fall below detectable levels. From the system of equations (3.2), we can see that at the pre-treatment steady state

$$\frac{1}{\delta}(1 - n_{rt})kV_I T_0 = T^*.$$

Therefore,

$$\begin{aligned} 0 &= (1 - n_p)N\delta T^* - cV_I, \\ &= (1 - n_p)N\delta\frac{1}{\delta}(1 - n_{rt})kV_I T_0 - cV_I, \\ &= (1 - n_p)(1 - n_{rt})kNT_0 - c. \end{aligned} \tag{3.10}$$

From which we can conclude that

$$c = (1 - n_c)kNT_0.$$

At the pre-treatment steady state no drug therapy has been administered. Therefore,

$$n_c = 0 \quad \Rightarrow \quad NkT_0 = c,$$

i.e.,

$$1 - \frac{T_0}{T_{s1}} < n_c. \tag{3.11}$$

If the patient is only administered with one drug, namely if they are prescribed the inhibitor to block virus particles but not the RT-inhibitor, then $n_{rt} = 0$. Hence, (3.11) gives that

$$n_p > 1 - \frac{T_0}{T_{s1}}. \tag{3.12}$$

Using values Perelson and Nelson [6] found to be typical, such that the normal level of T-cell count in a uninfected person is around $1000/\mu L$, and a patient is diagnosed with AIDS when the level drops below $200/\mu L$, we find that $n_p > 0.8$. This means that, if the drug is only administered at such a late point in the development of the virus, the drug has to be very strong to have an effect. Meanwhile if we assume that $T_{s1} = 1000$ and $T_0 = 500$, then $n_p > 0.5$. Therefore, the earlier the drug treatment starts, the less strong the treatment has to be in order to be effective [6]. However, if we administer both drugs at the same time, we see the following relationship

$$N_c = 1 - (1 - n_p)(1 - n_{rt}) < 1 - \frac{T_0}{T_{s1}}.$$

If we now let $T_0 = 200$ and $T_{s1} = 1000$, then we only need $n_p = 0.55$ and $n_{rt} = 0.55$. Hence, by using combination drug therapy, less strong drug are needed to be effective in reducing the level of

virus particles. Next, we explore whether there exists an infected steady state. At a steady state the following is true

$$s + pT\left(1 - \frac{T}{T_{max}}\right) - d_T T - kV_I T = 0, \tag{3.13}$$

$$(1 - n_{rt})kV_I T - \delta T^* = 0, \tag{3.14}$$

$$(1 - n_p)N\delta T^* - cV_I = 0, \tag{3.15}$$

$$n_p N\delta T^* - cV_{NI} = 0. \tag{3.16}$$

Using equation (3.15) we can see that

$$T^* = \frac{cV_I}{(1 - n_p)N\delta}.$$

Substituting T^* into equation (3.14) gives

$$\begin{aligned} 0 &= (1 - n_{rt})kV_I T - \delta\frac{cV_I}{(1 - n_p)N\delta} \\ &= V_I\left((1 - n_{rt})kT - \frac{c}{(1 - n_p)N}\right) \\ &= (1 - n_{rt})kT - \frac{c}{(1 - n_p)N}. \end{aligned}$$

Therefore,

$$(1 - n_{rt})kT = \frac{c}{(1 - n_p)N},$$

from which we can conclude that

$$\begin{aligned} T_{s2} &= \frac{c}{(1 - n_{rt})(1 - n_p)Nk} \\ &= \frac{c}{(1 - n_c)Nk}. \end{aligned} \tag{3.17}$$

Now that we have the infected steady state T count we can substitute this into equation (3.14) to find $\hat{V}_I$, the V_I at the infected steady state. We have

$$s + pT_{s2}\left(1 - \frac{T_{s2}}{T_{max}}\right) - d_T T_{s2} - kV_I T_{s2} = 0,$$

i.e.,

$$s + pT_{s2}\left(1 - \frac{T_{s2}}{T_{max}}\right) - d_T T_{s2} = kV_I T_{s2}.$$

Thus,

$$\frac{s}{kT_{s2}} + \frac{1}{k}\left(p\left[1 - \frac{T_{s2}}{T_{max}}\right] - d_T\right) = \hat{V}_I. \tag{3.18}$$

Using (3.18) we can find the value for T^* at the infected steady state $\hat{T}^*$. From (3.15) we get that

$$\hat{T}^* = \frac{c\hat{V}_I}{(1 - n_p)N\delta}.$$

The last value we need to evaluate is $\hat{V}_{NI}$. Substituting $\hat{T}^*$ into (3.16) we find

$$\hat{V}_{NI} = \frac{n_p N \delta T^*}{c} = \frac{n_p V_I}{1 - n_p}.$$

As before $n_c = 1 - (1 - n_{rt})(1 - n_p)$, therefore in absence of drugs $n_c = 0$, and when the drugs are less than perfect $0 < n_c < 1$. This infected steady state is only relevant if the virus exists, that is $\hat{V}_I > 0$, so

$$\begin{aligned} 0 &< \frac{s}{kT_{s2}} + \frac{1}{k}\left(p\left[1 - \frac{T_{s2}}{T_{max}}\right] - d_T\right) \\ &< \frac{s}{T_{s2}} + p - \frac{pT_{s2}}{T_{max}} - d_T \\ &< -\frac{pT_{s2}^2}{T_{max}} + T_{s2}(p - d_T) + s. \end{aligned} \tag{3.19}$$

However, if $V_I = 0$ then

$$\begin{aligned} T_{s2} &= \frac{-(p - d_T) \pm \sqrt{(p - d_T)^2 + 4\frac{ps}{T_{max}}}}{-\frac{2p}{T_{max}}} \\ &= \frac{T_{max}}{2p}\left[(p - d_T) + \sqrt{(p - d_T)^2 + 4\frac{ps}{T_{max}}}\right] = T_{s1}. \end{aligned}$$

Hence, at $V_I = 0$ the infected and uninfected steady states merge as expected due to the absence of the virus.Moreover, according to (3.19), we can see that as T_{s2} decreases, the left hand side of (3.19) increases. Therefore, when $V_I > 0$ it implies that $0 < T_{s2} < T_{s1}$. This is what we would expect to happen from the biological explanation of how the virus would behave. If we substitute our value for T_{s2} from equation (3.17) into equation (3.19) we have

$$\frac{(1 - n_c)Ns}{c} + \frac{1}{k}\left(p\left(1 - \frac{c}{T_{max}Nk(1 - n_c)}\right) - d_T\right) > 0.$$

Now in the case $s = 0$ the virus is no longer producing any viral particles,

$$\begin{aligned} \frac{1}{k}\left(p\left(1 - \frac{c}{T_{max}Nk(1 - n_c)}\right) - d_T\right) &> 0 \\ p\left(1 - \frac{c}{T_{max}Nk(1 - n_c)}\right) - d_T &> 0, \end{aligned}$$

as $d_T > 0$ we can assume that

$$p\left(1 - \frac{c}{T_{max}Nk(1 - n_c)}\right) > 0,$$

which implies

$$\frac{c}{T_{max}(1 - n_c)} < Nk \quad \Rightarrow V_I > 0.$$

Therefore, the following is also true

$$\frac{c}{T_{max}(1-n_c)} > Nk \quad \Rightarrow V_I < 0.$$

If we analyse the stability at this steady state, where there is no new production of viral particles, we will be able to see if the virus tends to such a position. If we let $\hat{T} = T_{s2}$, and $s = 0$, we can calculate the Jacobian J matrix of the system (3.2),

$$J = \begin{pmatrix} p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I & 0 & -k\hat{T} & 0 \\ (1-n_{rt})k\hat{V}_I & -\delta & (1-n_{rt})k\hat{T} & 0 \\ 0 & (1-n_p)N\delta & -c & 0 \\ 0 & n_pN\delta & 0 & -c \end{pmatrix}.$$

Therefore,

$$\begin{aligned} |J-\lambda I| = 0 \quad \Rightarrow 0 &= \begin{vmatrix} p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I - \lambda & 0 & -k\hat{T} & 0 \\ (1-n_{rt})k\hat{V}_I & -\delta-\lambda & (1-n_{rt})k\hat{T} & 0 \\ 0 & (1-n_p)N\delta & -c-\lambda & 0 \\ 0 & n_pN\delta & 0 & -c-\lambda \end{vmatrix} \\ &= (-c-\lambda)\begin{vmatrix} p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I - \lambda & 0 & -k\hat{T} \\ (1-n_{rt})k\hat{V}_I & -\delta-\lambda & (1-n_{rt})k\hat{T} \\ 0 & (1-n_p)N\delta & -c-\lambda \end{vmatrix}. \end{aligned} \tag{3.20}$$

Hence, we can see that the first eigenvalue is $\lambda_1 = -c$. Calculating the other eigenvalues, we get

$$\begin{vmatrix} p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I - \lambda & 0 & -k\hat{T} \\ (1-n_{rt})k\hat{V}_I & -\delta-\lambda & (1-n_{rt})k\hat{T} \\ 0 & (1-n_p)N\delta & -c-\lambda \end{vmatrix} = 0.$$

Calculating the determinant produces

$$\begin{aligned} &\left[p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I - \lambda\right]\left[(-\delta-\lambda)(-c-\lambda) - (1-n_p)N\delta(1-n_{rt})k\hat{T}\right] \\ &\quad - (1-n_{rt})k\hat{V}_I[k\hat{T}\delta N(1-n_p)] = 0, \end{aligned}$$

which implies

$$\begin{aligned} &\left[p\left(1-\frac{2\hat{T}}{T_{max}}\right) - d_T - k\hat{V}_I - \lambda\right]\left[(-\delta-\lambda)(-c-\lambda) - (1-n_c)\delta Nk\hat{T}\right] \\ &\quad - kV_Ik\hat{T}\delta N(1-n_c) = 0. \end{aligned} \tag{3.21}$$

By using the steady state value for $\hat{T}$, that is

$$\hat{T} = \frac{c}{Nk(1-n_c)}, \tag{3.22}$$

it is possible to simplify (3.21), so that

$$\begin{aligned}\left[p\left(1-\frac{2\hat{T}}{T_{max}}\right)-d_T-k\hat{V}_I-\lambda\right]\left[(-\delta-\lambda)(-c-\lambda)-\frac{(1-n_c)\delta Nkc}{Nk(1-n_c)}\right]\\ -\frac{k\hat{V}_Ik\delta Nc(1-n_c)}{Nk(1-n_c)}=0,\end{aligned} \tag{3.23}$$

i.e.,

$$\begin{aligned}0&=\left[p\left(1-\frac{2\hat{T}}{T_{max}}\right)-d_T-k\hat{V}_I-\lambda\right][(-\delta-\lambda)(-c-\lambda)-\delta c]-k\hat{V}_I\delta c\\ &=\left[p\left(1-\frac{2\hat{T}}{T_{max}}\right)-d_T-k\hat{V}_I-\lambda\right]\left[\lambda^2+(\delta+c)\lambda\right]\quad -k\hat{V}_I\delta c.\end{aligned} \tag{3.24}$$

Since equation (3.24) is of the form $\lambda^3+A\lambda^2+B\lambda+C=0$, where

$$\begin{aligned}A&=\delta+c+\frac{2p\hat{T}}{T_{max}}-(p-d_T)+kV_I,\\ B&=(\delta+c)\left[\frac{2p\hat{T}}{T_{max}}-(p-d_T)+kV_I\right],\quad C=c\delta kV_I.\end{aligned}$$

For stability purposes we do not need to evaluate the exact values for the eigenvalues, we just need to determine whether or not they are positive or negative. For the steady state to be stable we require the eigenvalues to have a negative real part. Routh-Hurwitz stability criterion shows that if $A>0$,$C>0$ and $AB-C>0$, then the eigenvalues do have a negative real part [11]. We can tell from an initial inspection that $C>0$ as it is made up of three constants which are all greater than zero and V_I which is the number of infectious viral particles that cannot be negative. From equation (3.13), we can deduce that

$$s+p\hat{T}\left(1-\frac{\hat{T}}{T_{max}}\right)-d_T\hat{T}=k\hat{V}_I\hat{T},$$

as $s=0$ it implies that

$$\begin{aligned}p\hat{T}-\frac{p\hat{T}^2}{T_{max}}-d_T\hat{T}&<k\hat{V}_I\hat{T}\\ (p-d_T)&<k\hat{V}_I+\frac{p\hat{T}}{T_{max}}.\end{aligned} \tag{3.25}$$

Hence, we can see from equation (3.25) that $A>0$. Next, we need to prove the last Routh-Hurwitz condition, that $AB-C>0$. Therefore,

$$A=\delta+c+\frac{2p\hat{T}}{T_{max}}-(p-d_T)+k\hat{V}_I,$$

by letting

$$B_1 = \frac{2p\hat{T}}{T_{max}} - (p - d_T) + k\hat{V}_I,$$

we get that

$$A = \delta + c + B_1, \quad B = (\delta + c)B_1.$$

From which we can conclude that

$$\begin{aligned} AB &= (\delta + c)^2 B_1 + (\delta + c)B_1^2 \\ &= (\delta + c)^2 \left[\frac{2p\hat{T}}{T_{max}} - (p - d_T) + k\hat{V}_I\right] + (\delta + c)\left[\frac{2p\hat{T}}{T_{max}} - (p - d_T) + k\hat{V}_I\right]^2 \\ &= (\delta^2 + 2\delta c + c^2)\left[\frac{2p\hat{T}}{T_{max}} - (p - d_T) + k\hat{V}_I\right] + (\delta + c)B_1^2 > 2\delta ck\hat{V}_I > \delta ck\hat{V}_I = C, \end{aligned}$$

which implies

$$AB > C \quad \Rightarrow AB - C > 0.$$

Therefore, if the steady state exists then it is classified as a stable steady point. As we showed earlier $T_{s2} < T_{s1}$, if T_{s2} exists. We also deduced from (3.9) that

$$1 - \frac{c}{NkT_{s1}} < n_c. \tag{3.26}$$

Therefore, if the stability condition (3.26) is not satisfied, then the only non-negative steady state is T_{s1}, hence there is no infected steady state. However, if (3.26) holds true, then there exists an infected stable steady state at T_{s2} . Therefore, we can deduce that there is a transcritical bifurcation point when $c = NkT_{s1}(1 - n_c)$.

Chapter 4

Numerical Approximations of Epidemic Models

Throughout the previous chapters, we have used computer algorithms to accurately predict solutions for the different population models, with specified initial conditions. For models where we cannot obtain explicit analytic formulas which show the trajectories as elementary functions, we can implement numerical methods that compute an approximate solution [3]. For example, when Kermack and McKerndrick solved the SIR model to find an explicit solution [9], they used the Taylor expansion when expanding the $e^{-\frac{R}{p}}$ term. When applying the exact solution shown in (1.25) to our Bombay epidemic problem, the size of recovered population reaches $20,000$. The Taylor expansion is not accurate for such high values of R as shown in Figure 4.1.

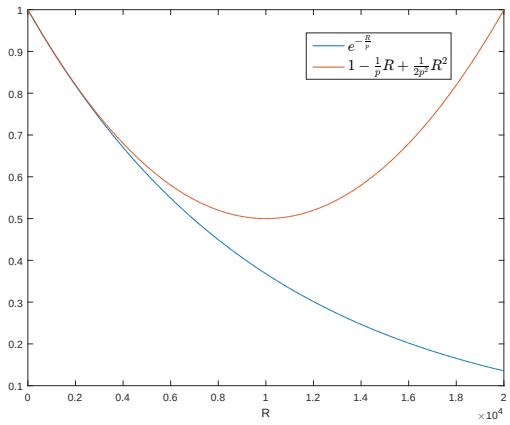

Figure 4.1: Comparison between $e^{-\frac{R}{p}}$ and the third term Taylor expansion.

Therefore, the exact solution will not be accurate when the size of the recovered population gets to large. If we graph the exact solution against an inbuilt Matlab solver approximation we see this relation. Figure 4.2, shows how the exact solution is fine until the Recovered population exceeds around 6000. This is at as similar value to when the Taylor expansion is no longer accurate. As the recovered population is also a parameter in the explicit solution for the infected and susceptible populations, the Taylor expansion error effects all three solutions.

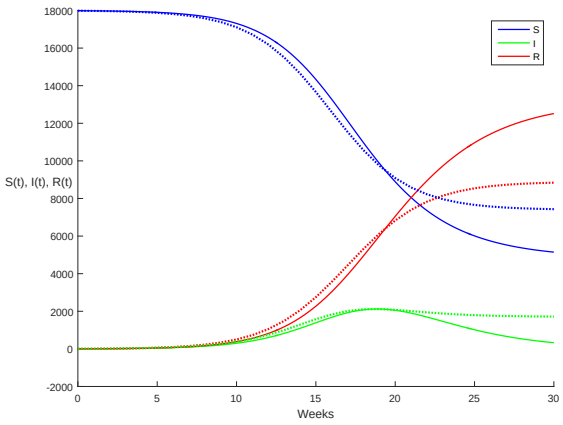

Figure 4.2: Comparison between explicit solution and the Ode45 approximation of the SIR model. The explicit solution is shown as the dotted lines.

Due to the error caused by the Taylor expansion in the derivation of the explicit solution, there are a variety of numerical approximations we can use to obtain a more accurate solution to the SIR model. In this chapter we will discuss a number of these, whilst investigating shortcomings associated with each method.

1 Explicit Euler Approximation

One of the simplest numerical approximations to calculate is the explicit Euler method. Let's begin with the following initial value problem,

$$y'(t) = f(t, y), \quad y(t_0) = y_0, \tag{4.1}$$

where $f(t, y)$ is a known function, and we want to solve (4.1) in the time interval $[t_0, t_f]$. Since the initial value is known, it is possible to approximate $y(t_0 + h)$ where h is the step size we want to use, using $y_0 + hf(t_0, y_0)$ [3]. Therefore if we let $y_1 = y_0 + hf(t_0, y_0)$, then we can use the same method to calculate y_2, as $y_2 = y_1 + hf(t_1, y_1)$ where $t_2 = t_1 + h$. Thus we can continue using this iteration until we get to the final point in our time interval t_f. This approximation technique is the explicit Euler method. We can therefore display the following explicit Euler formula for approximating y_{n+1},

$$y_{n+1} = y_n + hf(t_n, y_n), \tag{4.2}$$

where h is the step size of the the approximation. Figure 4.3 illustrates how the explicit Euler method approximates the solution, we can deduce that if the gradient is close to one then the approximation will be more accurate than when the gradient is much larger, where a smaller step size will be required to produce an accurate approximation. Since the explicit Euler method is a first order approximation, if we reduce the step size by a half we will in turn reduce the error by a half.

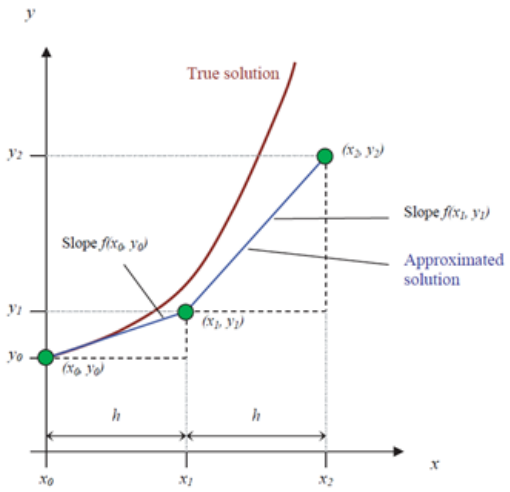

Figure 4.3: Schematic illustration of the explicit Euler method [1].

We can use the explicit Euler method to approximate the solution to SIR model for the Bombay plague epidemic discussed in chapter 1. Using the initial values and parameters Kermack and McKendrick [9] suggested, we can approximate the solution.

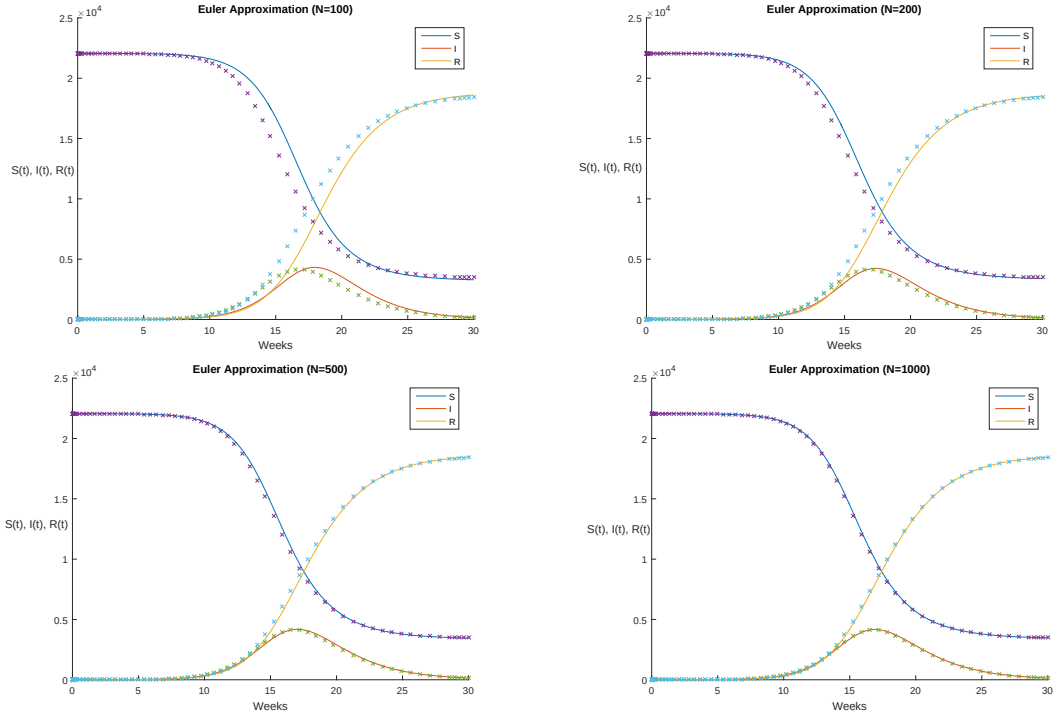

Figure 4.4: Euler approximations denoted by the crosses and Ode45 approximations denoted by the solid lines, of the Bombay plague epidemic SIR model using different number of steps.

Figure 4.4 shows that as we increase the number of steps of Euler uses, the approximation tends to the Ode45 solution.

2 Classical Fourth Order Runge Kutta Approximation

Although the explicit Euler method may produce a good approximation, we may require a more accurate approximation. A fourth order approximation will converge to the exact solution much quicker than a first order approximation. Higher order methods will produce more accurate approximations, however such methods require longer calculations. Therefore when deciding which calculation to use, we must balance the accuracy of the approximation, against the length of time required for the computation. The classical fourth order Runge Kutta approximation measures the slope at the mid-point and end points of the interval, a weighted average is then taken where more weight is placed on the slope at the mid point. Thus to approximate the solution of the initial value problem,

$$y'(t) = f(t, y), \qquad y(t_0) = y_0 \tag{4.3}$$

the following classical fourth order Runge Kutta method can be used,

$$\begin{aligned} K_1 &= hf(t, y_n), \\ K_2 &= hf(t + 0.5h, y_n + 0.5K_1), \\ K_3 &= hf(t + 0.5h, y_n + 0.5K_2), \\ K_4 &= hf(t + h, y_n + K_3), \\ y_{n+1} &= y_n + \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4). \end{aligned} \tag{4.4}$$

Inputting the same parameter values as we did earlier, when approximating the solution to the SIR model for the Bombay plague epidemic, we can use the classical fourth order Runge Kutta approximation to produce Figure 4.5.

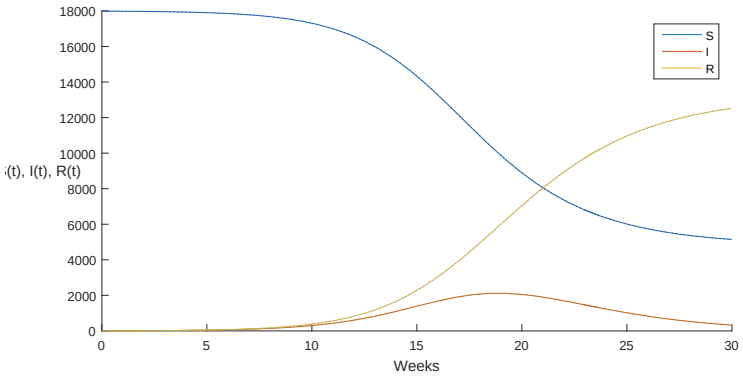

Figure 4.5: Approximation of the SIR Bombay epidemic model using the classical fourth order Runge Kutta method.

To deduce which approximation is more accurate we compare the errors of the corresponding approximations. As we do not have the exact solution of the SIR model, we will use the inbuilt Ode45 Matlab solver at at a really low tolerance setting as our exact solution. Figure 4.6 shows that the error of the explicit Euler approximation is much greater than that of the Runge Kutta approximation.

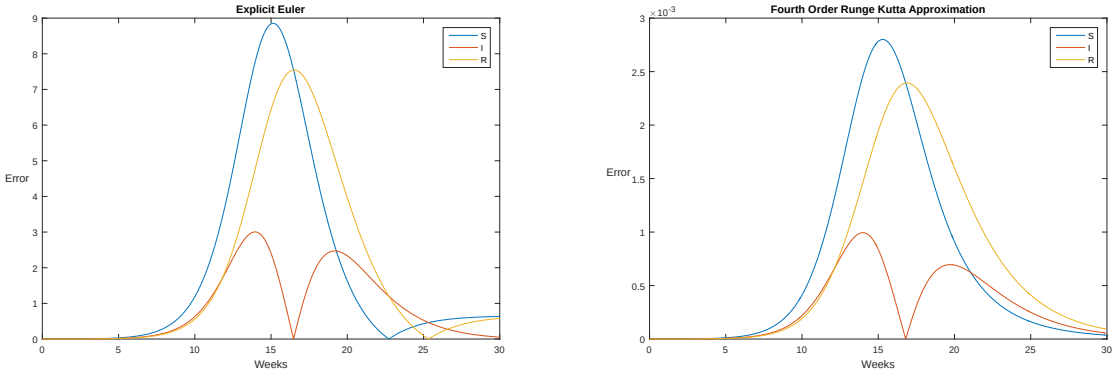

Figure 4.6: The error of the numerical approximations of the Bombay plague epidemic SIR model using a step size of $h = 0.001$ for the explicit Euler method and $h = 0.1$ for the Runge Kutta method.

However, as we discussed earlier we have to balance this increased accuracy against how long it takes to compute the approximation. Table 4.1 shows that the Runge Kutta method takes considerably longer to compute in comparison with the explicit Euler method. To produce an approximation with a step size of 0.001, the Runge Kutta method takes 5.7 times longer than the explicit Euler method. Therefore, although we getter a more accurate approximation using a higher order method, it takes considerably longer to produce. Thus, when deciding which method to use, we must deduce the preferred combination of accuracy and computational time.

	Time taken to compute the approximations (seconds)	
Step Size	Euler	Fourth order Runge Kutta
0.1	0.177	0.194
0.01	0.203	0.265
0.001	0.402	2.304
0.0001	2.504	19.625

Table 4.1: Time taken to compute the approximate solution to the SIR model at varying step sizes.

We can use the same Runge Kutta formula to approximate the solution to the HIV population model discussed in chapter two. Figure 4.7 shows the relation between the different population classes, it demonstrates how similar the classical Runge Kutta method is to the Matlab ode45 inbuilt solver.

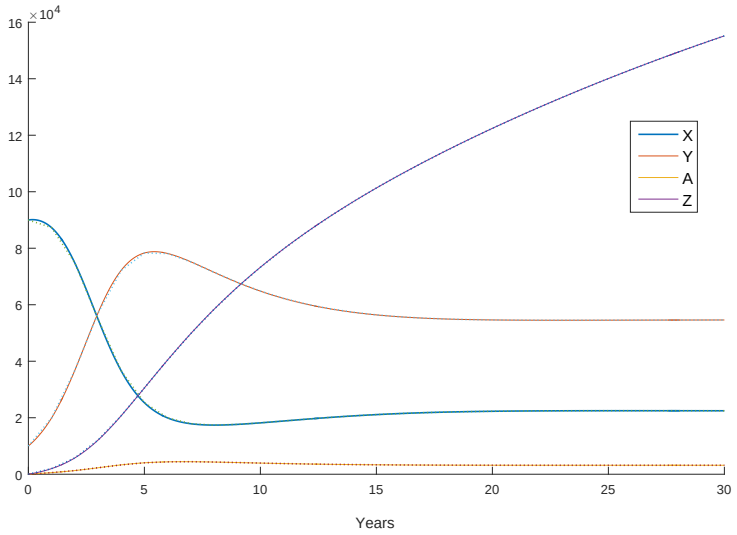

Figure 4.7: Approximation of the HIV population epidemic model using the classical fourth order Runge Kutta method, represented by the dotted lines, and by using the ode45 inbuilt Matlab solver.

If we compare the Runge Kutta approximation with the Euler approximation of HIV model we can see that the Euler approximation is less accurate when there is a greater change in the solution.

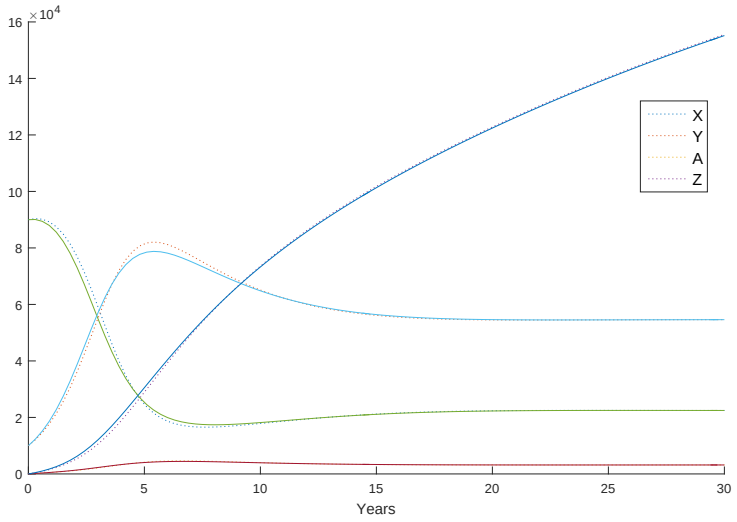

Figure 4.8: Approximation of the HIV population epidemic model using the classical fourth order Runge Kutta method and the explicit Euler method shown by the dotted lines.

3 Experimental Order of Accuracy

The Experimental Order of Accuracy (EOC) of an approximation shows how quickly the error decreases in response to a decrease in step size. If the exact solution is known then we can calculate

the error of the approximation, and thus the EOC is as follows,

$$\text{EOC} = \frac{\log\left|\frac{\epsilon_{n+1}}{\epsilon_n}\right|}{\log\left|\frac{h_{n+1}}{h_n}\right|} \tag{4.5}$$

where ϵ_n is the error of the approximation at the n^{th} step. If we decrease the step size by a half, a first order approximation should have an EOC of one, a second order approximation an EOC of two and so on. If we do not know the exact solution then Olof Runborg in his paper "Numerical Solutions of Differential Equations" [16], suggests two alternative methods. The first is to calculate the approximation at a very small step size and then treat this as the exact solution, this will enable us to use equation (4.5) to find the EOC. The second method is to compare the solution at varying step sizes. If we let $\hat{y}_h$ be the approximate solution with a step size of h, then we can formulate the EOC as follows,

$$\text{EOC} = \log_2\left|\frac{\hat{y}_h - \hat{y}_{\frac{h}{2}}}{\hat{y}_{\frac{h}{2}} - \hat{y}_{\frac{h}{4}}}\right|. \tag{4.6}$$

Using this formula we can now calculate the EOC of the explicit Euler and classical fourth order Runge Kutta approximations. We would expect the order of convergence to be approximately four for the Runge Kutta method and one for the forward Euler method. Applying the fourth order Runge Kutta approximation to the Bombay plague SIR model we see that as we decrease the step size, the EOC gets closer to four.

step size (h)	R(30)	$\hat{R}_h - \hat{R}_{\frac{h}{2}}$	$\frac{\hat{R}_h - \hat{R}_{\frac{h}{2}}}{\hat{R}_{\frac{h}{2}} - \hat{R}_{\frac{h}{4}}}$	$\log_2\left(\frac{\hat{R}_h - \hat{R}_{\frac{h}{2}}}{\hat{R}_{\frac{h}{2}} - \hat{R}_{\frac{h}{4}}}\right)$
0.3	$1.251775186308588 \times 10^4$	-2.74×10^{-3}	1.54E+01	3.940657
0.15	$1.251775460670866 \times 10^4$	-1.79×10^{-4}	1.57E+01	3.970058
0.075	$1.251775478538557 \times 10^4$	-1.14×10^{-5}	1.58E+01	3.984883
0.0375	$1.251775479678707 \times 10^4$	-7.20×10^{-7}	1.60E+01	3.996997
0.01875	$1.251775479750717 \times 10^4$	-4.51×10^{-8}	1.61E+01	4.008989
0.009375	$1.251775479755228 \times 10^4$	-2.80×10^{-9}		
0.004688	$1.251775479755508 \times 10^4$			

Table 4.2: This table uses the method shown in (4.6) to calculate the EOC values for the Runge Kutta method, using the approximated values of the Recovered population after 30 weeks.

Table 4.2 shows the EOC values at a specific point in time. Using Matlab we can calculate the EOC at every time point in the approximation. Figure 4.9 shows the results of these calculations, it is apparent that the EOC converges to four as we reduce the step size at every time point.

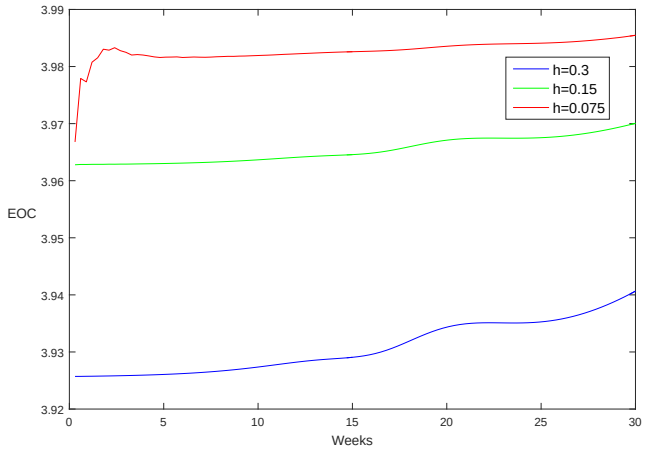

Figure 4.9: EOC of the Runge Kutta approximation of the SIR Bombay epidemic model at varying step sizes (h).

4 Adaptivity

Adapting the step size so that it is very small when we have a large change in the gradient of the solution, whilst using a larger step when there is less change, will enable us to produce a more efficient approximation. We will now investigate how adaptive step size controls work in the Ode45 solver and how it affects the solution. In 1969 Erwin Fehlberg introduced a Runge Kutta approximation that applied six function evaluations per step, this method was used by F.Shampine and his colleague H.A.Buddy to create the ODE solver RKF45. The formula for this method is illustrated below,

$$
\begin{aligned}
K_1 &= hf(t_i, y_i) \\
K_2 &= hf\left(t_i + \frac{h}{4}, y_i + \frac{1}{4}K1\right) \\
K_3 &= hf\left(t_i + \frac{3}{8}h, y_i + \frac{3}{32}K_1 + \frac{9}{32}K_2\right) \\
K_4 &= hf\left(t_i + \frac{12}{13}h, y_i + \frac{1932}{2197}K_1 - \frac{7200}{2197}K_2 + \frac{7296}{2197}K_3\right) \\
K_5 &= hf\left(t_i + h, y_i + \frac{439}{216}K_1 - 8K_2 + \frac{3680}{513}K_3 - \frac{845}{4104}K_4\right) \\
K_6 &= hf\left(t_i + \frac{h}{2}, y_i - \frac{8}{27}K_1 + 2K_2 - \frac{3544}{2565}K_3 + \frac{1859}{4104}K_4 - \frac{11}{40}K_5\right)
\end{aligned}
$$

where the fifth order approximation is given by,

$$z_{i+1} = z_i + \frac{25}{216}K_1 + \frac{1408}{2565}K_3 + \frac{2197}{4104}K_4 - \frac{1}{5}K_5$$

and the fourth order approximation as,

$$y_{i+1} = y_i + \frac{16}{135}K_1 + \frac{6656}{12825}K_3 + \frac{28561}{56430}K_4 - \frac{9}{50}K_5 + \frac{2}{55}K_6$$

This algorithm was the base for the creation of the ODE45 Matlab solver, this method was used for fifteen years until it was updated to a more efficient algorithm. The method which replaced it was the Dormand and Prince approximation, which uses a pair of Runge Kutta formulas derived for local extrapolation. The method developed by Dormand and Prince also uses six function evaluations to produce a fourth and fifth order Runge Kutta approximation solutions, as shown below,

$$\begin{aligned}
K_1 &= hf(t_i, y_i) \\
K_2 &= hf\left(t_i + \frac{h}{5}, y_i + \frac{1}{5}K1\right) \\
K_3 &= hf\left(t_i + \frac{3}{10}h, y_i + \frac{3}{40}K_1 + \frac{9}{40}K_2\right) \\
K_4 &= hf\left(t_i + \frac{4}{5}h, y_i + \frac{44}{45}K_1 - \frac{56}{15}K_2 + \frac{32}{9}K_3\right) \\
K_5 &= hf\left(t_i + \frac{8}{9}h, y_i + \frac{19372}{6561}K_1 - \frac{25360}{2187}K_2 + \frac{64448}{6561}K_3 - \frac{212}{729}K_4\right) \\
K_6 &= hf\left(t_i + h, y_i + \frac{9017}{3168}K_1 - \frac{355}{33}K_2 + \frac{46732}{5247}K_3 + \frac{49}{176}K_4 - \frac{5103}{18656}K_5\right) \\
K_7 &= hf\left(t_i + h, y_i + \frac{35}{384}K_1 + \frac{500}{1113}K_3 + \frac{125}{192}K_4 - \frac{2187}{6784}K_5 + \frac{11}{84}K_6\right)
\end{aligned}$$

where the fifth order approximation is given by,

$$z_{i+1} = z_i + \frac{35}{384}K_1 + \frac{500}{1113}K_3 + \frac{125}{192}K_4 - \frac{2187}{6784}K_5 + \frac{11}{84}K_6$$

and the fourth order approximation as,

$$y_{i+1} = y_i + \frac{5179}{57600}K_1 + \frac{7571}{16695}K_3 + \frac{393}{640}K_4 - \frac{92097}{339200}K_5 + \frac{187}{2100}K_6 + \frac{1}{40}K_7$$

The Matlab Ode45 solver uses an adaptive step size, it calculates the difference between the fourth and fifth order approximation which is said to be the error of the fourth order approximation. This error is then used to determine the step size used for the approximation at the next point. Although the Dormand and Prince method has seven stages, it only uses only six function evaluations as the first stage is the same as the last stage for the previous step. They choose the coefficients to minimize the fifth order approximation error, this is the main difference to the Fehlberg method which opted to minimize the fourth order approximation error. The Dormand and Prince method is a local extrapolation method as the higher order solution is used to continue the approximation at the next step. Calculating the experimental order of convergence of the fourth order approximation at different step sizes confirms that this Dormand and Prince method is of order four.

Step Size	Dormand and Prince Fourth Order Approximation	EOC
1	$1.67523957168987\times10^{-4}$	4.069
0.5	$1.67523569650797\times10^{-4}$	4.089
0.25	$1.67523546566328\times10^{-4}$	4.064
0.125	$1.67523545209468\times10^{-4}$	4.038
0.0625	$1.67523545128365\times10^{-4}$	4.021

Table 4.3: Dormand and Prince fourth order approximation of the recovered population at t=30 weeks, along with the corresponding order of convergence values.

If we then calculate the EOC of the fifth order Dormand and Prince method we see that this also confirms it as a fifth order approximation.

Step Size	Dormand and Prince Fifth Order Approximation	EOC
1	$1.675236787616821\times10^{-4}$	4.6396
0.5	$1.675235504595664\times10^{-4}$	4.813
0.25	$1.675235453125618\times10^{-4}$	4.856
0.125	$1.675235451294154\times10^{-4}$	5.007

Table 4.4: Dormand and Prince fifth order approximation of the recovered population at t=30 weeks, along with the corresponding order of convergence values.

By calculating the Truncation error coefficients we can compare the efficiency of the different approximations, where "The local truncation error at any time point in the integration interval is the numerical error due to the truncation of terms, if one assumes that the previous solution used to calculate the solution at the next time point is exact" [10]. F.Shampine proposed in the paper "Some Practical Runge-Kutta Formulas" [17], that we by implementing the following formula we could establish the local truncation error coefficients,

$$||T_{p+1}||_2 = \left(\sum_{j=1}^{\lambda_{p+1}} T_{p+1}^2\right)^{\frac{1}{2}}$$

where p is the order of the approximation. If we want to compare the relative efficiency of two methods say A and B, then Shampine proposed calculating the following computable relation will enable us to compare efficiency,

$$\frac{C_B}{C_A}\left(\frac{||T_{p+1}^B||}{||T_{p+1}^A||}\right)^{\frac{1}{p}} \tag{4.7}$$

where C_A is the amount of evaluations method A takes. By implementing these two formulas she deduced that the Dormand and Prince method was 11% more efficient than the Fehlberg method. Thus the improved Ode45 solver is more efficient due to replacing the Fehlberg formula with the

Dormand and Prince version.

When using the Ode45 solver in Matlab it is important to have the correct Tolerance settings, otherwise the level of accuracy of the approximation will be reduced. The relative tolerance is a measure of the error relative to the size of solution at any point in the approximation, it has a default setting of 1×10^{-3} for Ode45 in Matlab. Meanwhile the absolute tolerance is the threshold at point i in the approximation where the value of the solution is no longer important. Thus the absolute tolerance setting is very important for solutions close to zero and has a default setting of 1×10^{-6} for Ode45 in Matlab. The Ode45 solver calculates the local error as the difference between the fourth and fifth order approximations, this must be less than or equal to specified acceptable error which is a function of absolute tolerance and relative tolerance as shown below,

$$\text{tolerance}_j = \max(|y_j^k| \times \text{RelTol}, \text{AbsTol}_j).$$

We can see from the equation above that the relative tolerance is a scalar but the absolute tolerance is a vector, this enables us to set an absolute tolerance for each individual parameter. Therefore we can see that Ode45 adapts the step size depending on the relation between the pre-selected tolerance settings and the local error estimation. We can see from table 4.5, that when we reduce the absolute tolerance level the number of steps that ode45 takes increases, this arises from a smaller acceptable error size.

Absolute Tolerance Setting	Number of successful Steps ode45 Took
1×10^{-3}	85
1×10^{-6}	115
1×10^{-10}	151

Table 4.5: Comparison between the absolute tolerance setting and the number of steps the ode45 Matlab solver used when approximating the solution of the SIR epidemic model, whilst the relative tolerance was set to default.

We obtain a similar relation when we reduce the relative tolerance setting, however in comparison to the same reduction in absolute tolerance, the number of steps increases much quicker due to an decrease in relative tolerance. This is to be expected as all three of the parameters S,I and R do not spend a lot of time close to zero in the approximation.

Relative Tolerance Setting	Number of successful Steps ode45 Took
1×10^{-3}	115
1×10^{-6}	361
1×10^{-10}	1105

Table 4.6: Comparison between the relative tolerance setting and the number of steps the ode45 Matlab solver used when approximating the solution of the SIR epidemic model, whilst the absolute tolerance was set to default.

Comparing how ode45 compares against a different Nonstiff Matlab solver such as ode23, we see that to achieve the same relative tolerance error level, ode23 has to take many more steps than ode45. The ode23 Matlab solver uses a combination of second and third order Runge Kutta formulas in a similar fashion to how ode45 works [4].

Relative Tolerance Setting	ode45	ode23
1×10^{-3}	115	151
1×10^{-4}	181	307
1×10^{-5}	241	646
1×10^{-6}	361	1342
1×10^{-7}	529	2686
1×10^{-8}	763	5023

Table 4.7: Comparison between of the number of successful steps the ode45 and ode23 Matlab solvers used at different relative tolerance setting when approximating the solution of the SIR epidemic model, whilst the absolute tolerance was set to default.

If we now compare the ode45 solver which implements adaptivity, against the Dormand and Prince approximation with a step size of 0.25, we can see that there is not a lot of difference in the solutions, however the use of adaptivity by ode45 reduces the number of steps necessary to produce an accurate solution by only decreasing the step size when there is an unacceptable local error.

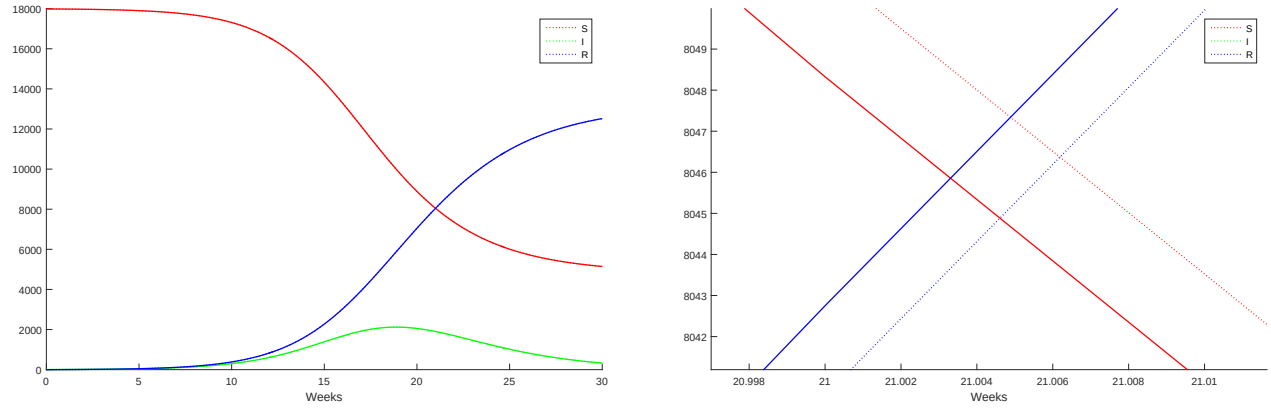

Figure 4.10: The solution to the SIR model using the Matlab inbuilt command function ode45, and using the Dormand and Prince fifth order Runge Kutta method with a step size of 0.25. The Dormand and Prince approximation is shown using the dotted lines. The graph on the right is a zoomed in version of the one on the left where the susceptible and recovered population lines intersect.

If we now look at the phase plane diagram of the SIR model during the Bombay plague epidemic we can see that through looking at different initial population levels, we can reproduce the phase plane diagram we expected to get using the theory shown in Figure 1.3.

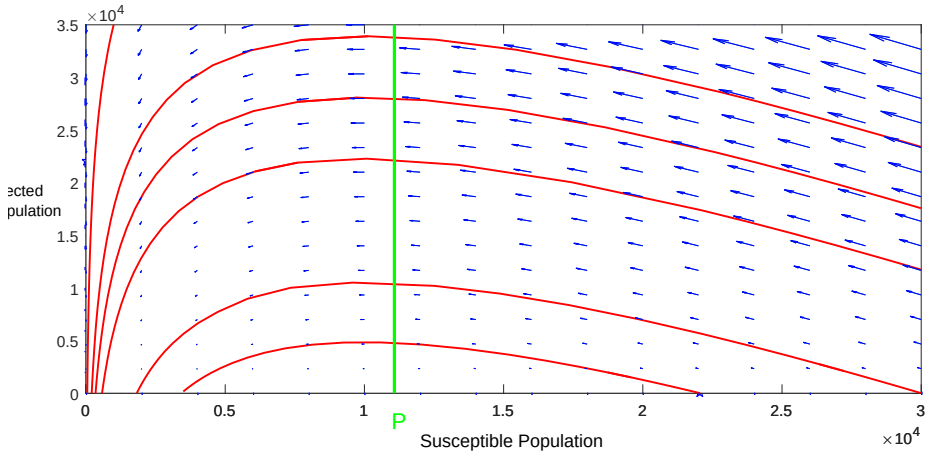

Figure 4.11: The phase plane diagram for SIR Bombay epidemic model using different initial conditions.

Therefore regardless of the initial conditions the Infected population will always tend to zero over time. We can also confirm a relation we discovered analyticity earlier, that the peak in the Infected population occurs when $S = p$, if $S_0 > p$. In this chapter we have shown how numerical approximations can provide accurate solutions to problems when we cannot obtain the explicit analytic solutions. Furthermore, we have been able to confirm expected behaviors and relations between different parameters. Thus, we have shown how important numerical approximations are when investigating population models.

Chapter 5

Conclusion

We began this investigation into mathematical modelling by explaining the process for selecting a suitable epidemiological model. To achieve this we had to identify the key parameters that effect the populations we intended to model. Once a model is selected we can compare the modelled predictions against the real data collected. In the case of our SIR model, Kermack and McKendrick compared the model against data collected from the 1906 Bombay epidemic [9]. They found that by carefully selecting the parameters, they could accurately model the change in the recovered population.

We then investigated the problem posed by diseases which are harder to model. An example of such a problem is modelling how HIV is spread through a population. Murray suggests the use of a epidemic model consisting of four population classes [12]. The proposed demographic model included realistic assumptions, such as there existed natural death in all four categories. We evaluated that such a model has two steady states, one uninfected steady state and one infected, both of which are stable. Hence it is possible that HIV will not naturally die off. Therefore it becomes important that there are effective treatments which can reduce the spread of HIV and prevent an epidemic occurring.

Thus, we investigated the biological make up of HIV and deduced that drug therapy was the most effective treatment. Perelson and Nelson created a suitable combination drug therapy model [6]. This modelled how certain drugs effect not only the level of virus particles, but also the production of T-cells, which are vital in the function of the immune system. We concluded that there existed an uninfected and infected steady state. Again we evaluated that both states were stable, where there existed a transcritical bifurcation point at $c = NkT_{s1}(1 - n_c)$. Therefore the viral clearance rate is highly dependent on the effectiveness of the drug therapy n_c. Earlier in this chapter we showed that administering more than one drug increases the effectiveness of the therapy and thus slows down the transmission of the virus. We can therefore conclude that to tackle HIV effectively, we require a combination of quick diagnosis and effective drug therapy.

For models such as the ones used in this paper we cannot find accurate analytical solutions. Thus numerical approximations become extremely important in the analysis of such models. Therefore we have investigated several numerical methods to deduce the advantages and disadvantages of such methods. We concluded that although higher order methods are more accurate, the increase in computational time has to be balanced against the need for such a high degree of accuracy. Therefore

we analysed adaptive numerical approximations to evaluate how they can improve efficiency. We focused on the ode45 Matlab solver which uses two Runge Kutta methods developed by Dormand and Prince. They used a fifth order approximation and compared the solution to one produced by a fourth order approximation, the difference between them is classed as the relative error. The step size is then adapted depending on the size of the relative error. Using this method we can produce accurate approximations using less steps than a fixed step size method. To conclude, we have shown how important mathematics can be in epidemiological modelling, we have been able to show that through a combination of model analysis and numerical approximations of the solutions, we can further the understanding of how epidemics occur, whilst finding ways to try and prevent them happening.

1 Future Work

One of the major issues with approximating the SIR model is we do not have an accurate explicit solution. This was due to the third term Taylor expansion of $e^{-\frac{R}{p}}$ not being accurate for large values of R. By using more terms in the Taylor expansion, it becomes more accurate. Figure 5.1, shows how the increase in terms oscillates whilst it converges to the actual solution.

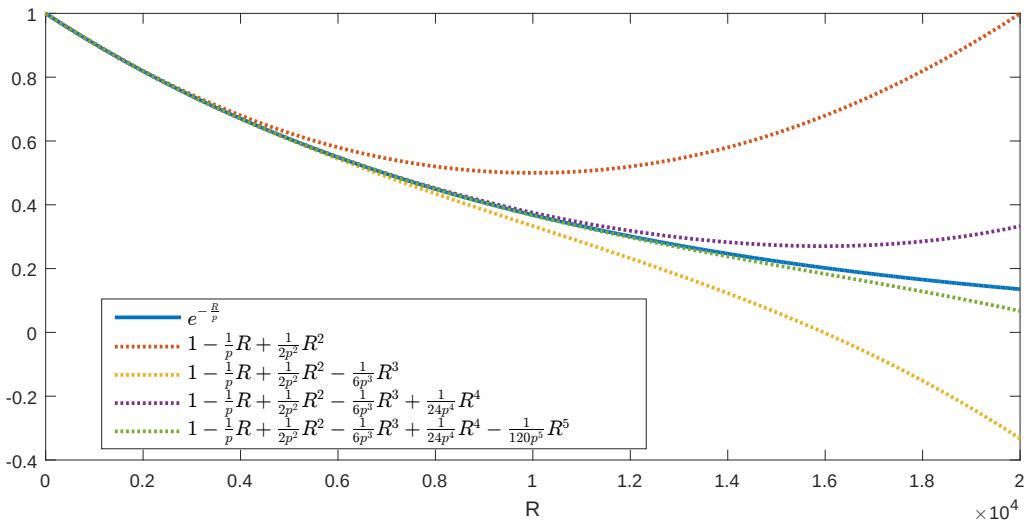

Figure 5.1: Comparison of the numbers of terms used in the Taylor expansion of $e^{-\frac{R}{p}}$.

By using more Taylor expansion terms we can calculate a more accurate explicit solution for $R(t)$. This will enable to improve our comparison of the different numerical approximations.

During this dissertation we used a Aids model proposed by Murray [6]. This model had certain limitations, such as it used one system of equations to model the entire homosexual population. This meant that parameters such as the incubation period are assumed to be the same for all males regardless of age. In reality we know that such assumptions are not realistic. Therefore we can

use a more complex model which accounts for such differences in parameters values. Griffihs states that "Recent research has shown that the incubation period of HIV, the virus that leads to AIDS, is clearly age dependent" [5]. He therefore proposes the use of an age dependent model, which also takes into account the effects of government sponsored AIDS prevention programmes. He created a systems of equations for each of the three different age groups. This enables us to identify which age groups should be targeted by campaigns to be most effective. We can also adapt such a model to understand how different age groups respond to treatment programmes. Applying the analysis in our chapter on drug therapy for HIV patients, to specific population groups may enable us to identify more effective age dependent treatments. The successful prevention of future epidemics depends on our ability to create effective models in order to identify those most at risk. This will enable governments to have targeted campaigns which will have higher efficiency in combating epidemics.

Appendices

Figure 1.1

```
function SIR =SIR(t,y)
r = 0.6;
a = 0.1;
SIR(1) =-r*y(1)*y(2);
SIR(2) = r*y(1)*y(2)-a*y(2);
SIR(3) = a*y(2);
SIR = [SIR(1) SIR(2) SIR(3)]';
ylabel('susceptible, infected, recovered')
\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%
clear
to = 0;
tf =30;
yo = [99 01 0];
[t y] = ode45('SIR',[to tf],yo);
plot(t,y(:,1),t,y(:,2),t,y(:,3))
legend('S','I','R')
title('SIR model')
xlabel('time')
```

Figure 1.3

```
clear
t=0:1:30;
g=890*((sech(0.2*t-3.4)).^2);

plot(t,g)
xlabel('Weeks')
ylabel('Deaths( \approx dR/dt)')
```

Figure 4.1

```
r=0:20000;
p=10002.59
y=exp(-(r)/p)

plot(r,y)
hold on
y1=1-((1/p).*r)+((1/(2*p.^2))*r.^2)

plot(r,y1)
xlabel('R')
```

Figure 4.2

```
function rk4_bombay(a, b, N, alpha)
format long
alpha=[17986.87154 6 0]
b=30
a=0
 m = size(alpha,1);

N1=50

h1 = (b-a)/N1;        %the step size
t1(1) = a;
w1(:,1) = alpha;     %initial conditions
t2(1) = a;
y(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-6,'AbsTol',1e-6,'NonNegative',[1 2 3]);
[t2 y1] = ode45('bombay',[a b],alpha, options);

r=0.00004998701123
a1=0.5   %this is a
p=10002.59843
g=0.8  %this is alpha
S=17986.87154  %this is S_0
i0=6   %this is I_0
 t11=0:1:30;
R1=((a1.^2)./(r.^2 .*S)).*((r.*S./a1)-1+g.*tanh(((0.2).*t11)-3.4));
S1=S.*(exp(-((R1)/p)));
```

```
I1=-S1+p.*log(S1)+i0+S-p.*log(S);
hold on
 plot(t2,y1)
legend('S','I','R')
 xlabel('Weeks')
ylabel('S(t), I(t), R(t)')
plot(t11,R1,':',t11,S1,':',t11,I1,':')

function bombay =f(t,y)
r=0.00004998701123;
a1=0.5;
bombay(1) =-r.*y(1).*y(2);
bombay(2) = r.*y(1).*y(2)-a1.*y(2);
bombay(3) = a1.*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';
```

Figure 4.4

```
function ie(a, b, N, alpha)
hax=axes;
alpha=[22058 1 0]
b=30
a=0
N=300
m = size(alpha,1);


h = (b-a)/N;        %the step size
t(1) = a;
w(:,1) = alpha;     %initial conditions

for i = 1:N
  t(i+1)=t(i)+h;
  w(:,i+1)=w(:,i)+h*f(t(i),w(:,i));
end
hold on
plot(t,w ,'LineWidth',2)
legend('S','I','R')
xlabel('Weeks')
ylabel('S(t), I(t), R(t)')
title('Euler Approximation')
axis([0,30,0,25000])

error=abs(1.675135451231053e4-w(3,N+1))
```

```
function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';

\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%

function test1()
 euler_systems(0, 30, 300, [22058 1 0]);
 end
```

Figure 4.5

```
function rk4_system(a, b, N, alpha)
format long
m = size(alpha,1);
tspan=[0:1:N];
y(:,1) = alpha;
h = (b-a)/N;        %the step size
t(1) = a;
w(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
[t1 y1] = ode45('bombay',tspan,alpha, options);



for i = 1:N


k1 = h*f(t(i), w(:,i));
   k2 = h*f(t(i)+h/2, w(:,i)+0.5*k1);
   k3 = h*f(t(i)+h/2, w(:,i)+0.5*k2);
   k4 = h*f(t(i)+h, w(:,i)+k3);
   w(:,i+1) = w(:,i) + (k1 + 2*k2 + 2*k3 + k4)/6;
   t(i+1) = t(i) + h;
```

```
end
length(y);
error=abs(y1-w');
plot(t,w,':')
legend('S','I','R')
title('Fourth order Runge Kutta')
xlabel('Weeks')
ylabel('Error')



function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';
\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%
 function test()
 rk4_system(0, 30, 300, [22058 1 0]);
end
```

Figure 4.6

```
function rk4_systems(a, b, N, alpha)
format long
alpha=[22058 1 0]
b=30
a=0
N=3000;
m = size(alpha,1);
h = (b-a)/N;
tspan=[0:h:30];
y(:,1) = alpha;
        %the step size
t(1) = a;
y(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
[t1 y1] = ode45('bombay',tspan,alpha, options);
```

```
for i = 1:N
     t(i+1)=t(i)+h;
  y(:,i+1)=y(:,i)+h*f(t(i),y(:,i));

end
length(y)
error=abs(y1-y');
plot(t,error)
legend('S','I','R')
xlabel('Weeks')
title('Forward Euler')
ylabel('Error')


function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';

\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%

function rk4_system(a, b, N, alpha)
format long
m = size(alpha,1);
tspan=[0:1:N];
y(:,1) = alpha;
h = (b-a)/N;        %the step size
t(1) = a;
w(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
[t1 y1] = ode45('bombay',tspan,alpha, options);

for i = 1:N

k1 = h*f(t(i), w(:,i));
   k2 = h*f(t(i)+h/2, w(:,i)+0.5*k1);
   k3 = h*f(t(i)+h/2, w(:,i)+0.5*k2);
   k4 = h*f(t(i)+h, w(:,i)+k3);
   w(:,i+1) = w(:,i) + (k1 + 2*k2 + 2*k3 + k4)/6;
   t(i+1) = t(i) + h;
end
length(y);
```

```
error=abs(y1-w');
plot(t,error)
legend('S','I','R')
title('Fourth order Runge Kutta')
xlabel('Weeks')
ylabel('Error')


function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';

\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%
 function test()
 rk4_system(0, 30, 300, [22058 1 0]);
end
 function test1()
euler_systems(0, 30, 300, [22058 1 0]);
 end
```

Figure 4.7

```
function rk4_aids(a, b, N, alpha)
format long
alpha = [90000 10000 0 0];
b=30
a=0
 m = size(alpha,1);


N1=100



h1 = (b-a)/N1;        %the step size
t1(1) = a;
w1(:,1) = alpha;     %initial conditions
t2(1) = a;
```

```
y(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-6,'AbsTol',1e-6,'NonNegative',[1 2 3]);
[t2 y1] = ode45('aids',[a b],alpha, options);


for i = 1:N1
   k11 = h1*f(t1(i), w1(:,i));
   k21 = h1*f(t1(i)+h1/2, w1(:,i)+0.5*k11);
   k31 = h1*f(t1(i)+h1/2, w1(:,i)+0.5*k21);
   k41 = h1*f(t1(i)+h1, w1(:,i)+k31);
   w1(:,i+1) = w1(:,i) + (k11 + 2*k21 + 2*k31 + k41)/6;
   t1(i+1) = t1(i) + h1;
end



hold on
 plot(t1,w1)
 legend('X','Y','A','Z')



function yaids =f(t,y)
B=13333.3
v=0.2
u=1/32
d=1
p=0.3
R=5.15
r=0.83
a=1
t=0:1:30;
b=0.515
c=2
N=100000


yaids(1) =B-u*y(1)-b*c*y(2)*y(1)/N;
yaids(2) = b*c*y(2)*y(1)/N-(v+u)*y(2);
yaids(3) = p*v*y(2)-(d+u)*y(3);
yaids(4)=(1-p)*v*y(2)-u*y(4);
yaids = [yaids(1) yaids(2) yaids(3) yaids(4)]';
```

Figure 4.8

```
function eu(a, b, N, alpha)

hax=axes;
alpha=[90000 10000 0 0];
m = size(alpha,1);
a=0
b=30
N=100

h = (b-a)/N;        %the step size
t(1) = a;
w(:,1) = alpha;     %initial conditions
t2(1) = a;
y(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-3,'NonNegative',[1 2 3]);
[t2 y1] = ode45('aids',[a b],alpha, options);

for i = 1:N
  t(i+1)=t(i)+h;
  w(:,i+1)=w(:,i)+h*f(t(i),w(:,i));
end




hold on
plot(t,w,':' )
legend('X','Y','A','Z')
xlabel('Weeks')
ylabel('S(t), I(t), R(t)')
title('Euler Approximation (N=1000)')




function yaids =f(t,y)
B=13333.3
v=0.2
u=1/32
```

```
d=1
p=0.3
R=5.15
r=0.83
a=1
t=0:1:30;
b=0.515
c=2
N=100000


yaids(1) =B-u*y(1)-b*c*y(2)*y(1)/N;
yaids(2) = b*c*y(2)*y(1)/N-(v+u)*y(2);
yaids(3) = p*v*y(2)-(d+u)*y(3);
yaids(4)=(1-p)*v*y(2)-u*y(4);
yaids = [yaids(1) yaids(2) yaids(3) yaids(4)]';
```

Figure 4.9

```
function rk4_bombay(a, b, N, alpha)
format long
alpha=[17986.87154 6 0]
b=30
a=0
 m = size(alpha,1);
N1=6400
h1 = (b-a)/N1;        %the step size
t1(1) = a;
w(:,1) = alpha;     %initial conditions
t2(1) = a;
y(:,1) = alpha;     %initial conditions
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
[t2 y1] = ode45('bombay',[a b],alpha, options);

fp1=fopen('rk4_output', 'w')
 fprintf(fp1,'%4.2f %12.10f %12.10f %12.10f\n', t1(1), w(1,1), w(2,1), w(3,1));
for i = 1:N1
   k11 = h1*f(t1(i), w(:,i));
   k21 = h1*f(t1(i)+h1/2, w(:,i)+0.5*k11);
   k31 = h1*f(t1(i)+h1/2, w(:,i)+0.5*k21);
   k41 = h1*f(t1(i)+h1, w(:,i)+k31);
   w(:,i+1) = w(:,i) + (k11 + 2*k21 + 2*k31 + k41)/6;
```

```
   t1(i+1) = t1(i) + h1;
  fprintf(fp1,'%4.2f %12.12f %12.12f %12.12f\n', t1(i+1), w(1,i+1), w(2,i+1), w(3,i+1));
 end
 fclose(fp1);

hold on
w(3,N1+1)
 plot(t2,y1,'r')

function bombay =f(t,y)
r=0.00004998701123;
a1=0.5;
bombay(1) =-r.*y(1).*y(2);
bombay(2) = r.*y(1).*y(2)-a1.*y(2);
bombay(3) = a1.*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';

\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%

function eoc_rk4()
 f1=load('rk4_output_100');
 f2=load('rk4_output_200');
 f3=load('rk4_output_400');
 f4=load('rk4_output_800');
 f5=load('rk4_output_1600');

 len = length(f1(:,1))
 j=1
 k=1
 p=1
 m=1

 for i=1:1:len

 f1(i,1), f2(j,1), f3(k,1),f4(p,1),f5(m,1)
     eoc1(i)=log2(abs(abs((f1(i,4))-abs(f2(j,4)))./abs((f2(j,4))-abs(f3(k,4)))))
     eoc2(i)=log2(abs((abs(f2(j,4))-abs(f3(k,4)))./(abs(f3(k,4))-abs(f4(p,4)))));
     eoc3(i)=log2(abs((f3(k,2)-f4(p,2))./(f4(p,2)-f5(m,2))))
     j=j+2;
     k=k+4;
     p=p+8;
     m=m+16;
 end
```

```
 hold off
g=f1(:,1);
 plot(g,eoc1,'b')
 hold on
  plot(g,eoc2,'g')
     hold on
   plot(g,eoc3,'r')
 xlabel('Weeks')
ylabel('EOC')
legend('h=0.3','h=0.15','h=0.075')
```

Figure 5.0

```
function rkdormand_systems(a, b, N, alpha)
format long
alpha=[22058 1 1]
b=30
a1=0
N=120
m = size(alpha,1);
h = (b-a1)/N;
t(1) = a1;
w(:,1) = alpha;
w1(:,1) = alpha;

for i = 1:N

   k1 = h*f(t(i), w(:,i));
   k2 = h*f(t(i)+h/5, w(:,i)+(1/5)*k1);
   k3 = h*f(t(i)+(3/10)*h, w(:,i)+(3/40)*k1+(9/40)*k2);
   k4 = h*f(t(i)+(4/5)*h, w(:,i)+(44/45)*k1-(56/15)*k2+(32/9)*k3);
   k5 = h*f(t(i)+(8/9)*h, w(:,i)+(19372/6561)*k1-(25360/2187)*k2
          +(64448/6561)*k3-(212/729)*k4);
   k6 = h*f(t(i)+h, w(:,i)+(9017/3168)*k1-(355/33)*k2+(46732/5247)*k3+(49/176)*k4
          -(5103/18656)*k5);
   k7 = h*f(t(i)+h, w(:,i)+(35/384)*k1+(500/1113)*k3+(125/192)*k4
           -(2187/6784)*k5+(11/84)*k6);


   w(:,i+1) = w(:,i)+ (35/384)*k1 + (500/1113)*k3 + (125/192)*k4 -(2187/6784)*k5
                    +(11/84)*k6 ;

   w1(:,i+1) = w1(:,i) + (5179/57600)*k1 + (7571/16695)*k3+(393/640)*k4
```

```
                    -(92097/339200)*k5+(187/2100)*k6 +(1/40)*k7;
   t(i+1) = t(i) + h;

end
hold on
 plot(t,w,':')
legend('S','I','R')
hold on
h = (b-a1)/N;
tspan=[0:h:30];
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
[t1 y1] = ode45('bombay',tspan,alpha, options);

plot(t,y1)

e=w(:,N+1)


function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';
```

Figure 5.1

```
 to = 0;
tf =30;
h=1;
hax=axes;
n=30/h;
yo = [22058 1 0];

tspan=0:5:30;
options=odeset('RelTol',1e-3,'AbsTol',1e-4,'NonNegative',[1 2 3]);
options =odeset('OutputFcn',@odephas2,'OutputSel',[1 2]);
[t1 y1 stats] = ode45('bombay',[to tf],yo, options)
hold on
[t2 y2 stats] = ode45('bombay',[to tf],[15000 7058 0], options)
```

```
hold on
[t3 y3 stats] = ode45('bombay',[to tf],[12000 10058 0], options)
hold on
[t4 y4 stats] = ode45('bombay',[to tf],[10000 12058 0], options)
hold on
[t5 y5 stats] = ode45('bombay',[to tf],[18000 4058 0], options)
hold on
[t6 y6 stats] = ode45('bombay',[to tf],[6000 16058 0], options)
xlabel('Susceptible Population')
ylabel('Infected Population')

hold on
q=[11076.94,11076.94]
d=[0,10982.06]
plot(q,d,'color','r')
hold on
line([0 22059],[22059 0]);
 plot([0 22059],[22059 0]);


\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%\%


function bombay =f(t,y)
r =0.000045139 ;
a = 0.5;
bombay(1) =-r*y(1)*y(2);
bombay(2) = r*y(1)*y(2)-a*y(2);
bombay(3) = a*y(2);
bombay = [bombay(1) bombay(2) bombay(3)]';
```

Bibliography

[1] S. Amen, P. Bilokon, A. Brinley-Codd, M. Fofaria, and T. Shah, *Numerical Solutions of Differential Equations*, Imperial College London, 2004.

[2] R.M. Anderson, G.F. Medley, May.R.M., and A.M. Johnson, *A Preliminary Study of the Transmission Dynamics of the Human Immunodeficiency Virus (HIV), the Cuasative Agent of Aids*, IMA Journal of Mathematics Applied in Medicineand Biology **3** (1986), 229-263.

[3] S.P. Ellner and J. Guckenheimer, *Dynamic Models in Biology*, Princeton University Press, 2006.

[4] J. Faires and R. Burden, *Numerical Methods*, Vol. 4, Cengage Learning, 2012.

[5] J. Griffiths, D. Lowrie, and J. Williams, *An Age Structured Model for the AIDS Epidemic*, European Journal of Operational Research **124** (2000), 1-14.

[6] T. Gross, J. Faull, S. Ketteridge, and D. Springham, *Introductory Microbiology*, Vol. 1, Chapman and Hall, 1995.

[7] D.D. Ho, A.U. Neumann, W. Chen, J.M. Leonard, and M. Markowitz, *Rapid Turnover of Plasma Virions and CD4 Lymphocytes in HIV-1 Infection*, Nature **373** (1995), 123-126.

[8] J.N. Kapur, *Mathematical Modeling*, Vol. 1, New Age International, 1988.

[9] W.O. Kermack and A.G. McKendrick, *A Contribution to the Mathematical Theory of Epidemics*, A, 1927.

[10] M.R. King and N.A. Mody, *Numerical and Statisical Methods for Bioengineering: Applications in Matlab*, 1, Cambridge University Press, 2010.

[11] S. Koskie, *Routh's Stability Criterion* **1** (2007), no. 1, 1-6.

[12] J.D. Murray, *Mathematical Biology*, 2, Springer-Verlag Berlin Heidelberg, 1989.

[13] Office for National Statistics, *Integrated Household Survey, January to December 2013: Experimental Statistics* (2014).

[14] A.S. Perelson and P.W. Nelson, *Mathematical Analysis of HIV-1 Dynamics in Vivo*, Vol. 41, Society of Industrial and Applied Mathematics, 1999.

[15] Peterman, Drotman, and Curran, *Data on HIV* (1985), 573-582.

[16] O. Rumborg, *Verifying Numerical Convergence Rates*, KTH Computer Science and Communication, 2012.

[17] L.F. Shampine, *Some Pratical Runge-Kutta Formulas*, Vol. 46, Mathematics of Computation, 1986.

[18] L.F. Shampine, G. Gladwelll, and S. Thompson, *Solving ODE's with Matlab*, Vol. 1, Cambridge University Press, 2008.

[19] UNAIDS, *Report on the global AIDS epidemic: Executive summary* (2006).

[20] UNAIDS, *Report on the Global HIV/AIDS epidemic* (2002).